LA-UR-23-23408
Approved for public release; distribution is unlimited.

Title: A Waveform Detection Tutorial: A (Very) Short Course

Author(s): Carmichael, Joshua Daniel

Intended for: Teaching post-doctoral students (knowledge transfer).
Report

Issued: 2023-04-03

Los Alamos
NATIONAL LABORATORY

A Waveform Detection Tutorial: A (Very) Short Course

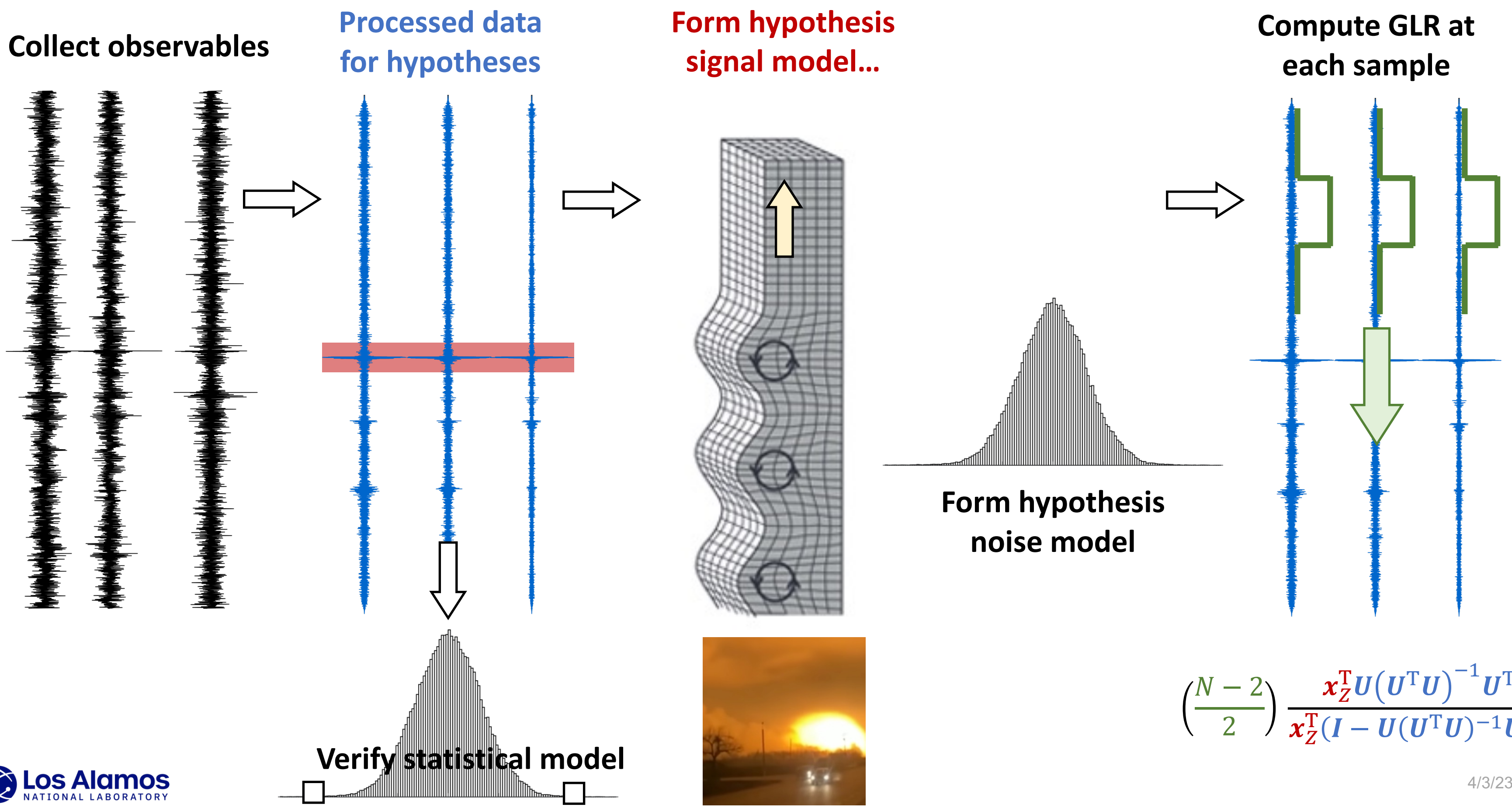
***OR**: A Los Alamos National Laboratory Geophysical Explosion Monitoring (GEM) Team Waveform Detection Tutorial*

02 March 2023

What’s All This Then?

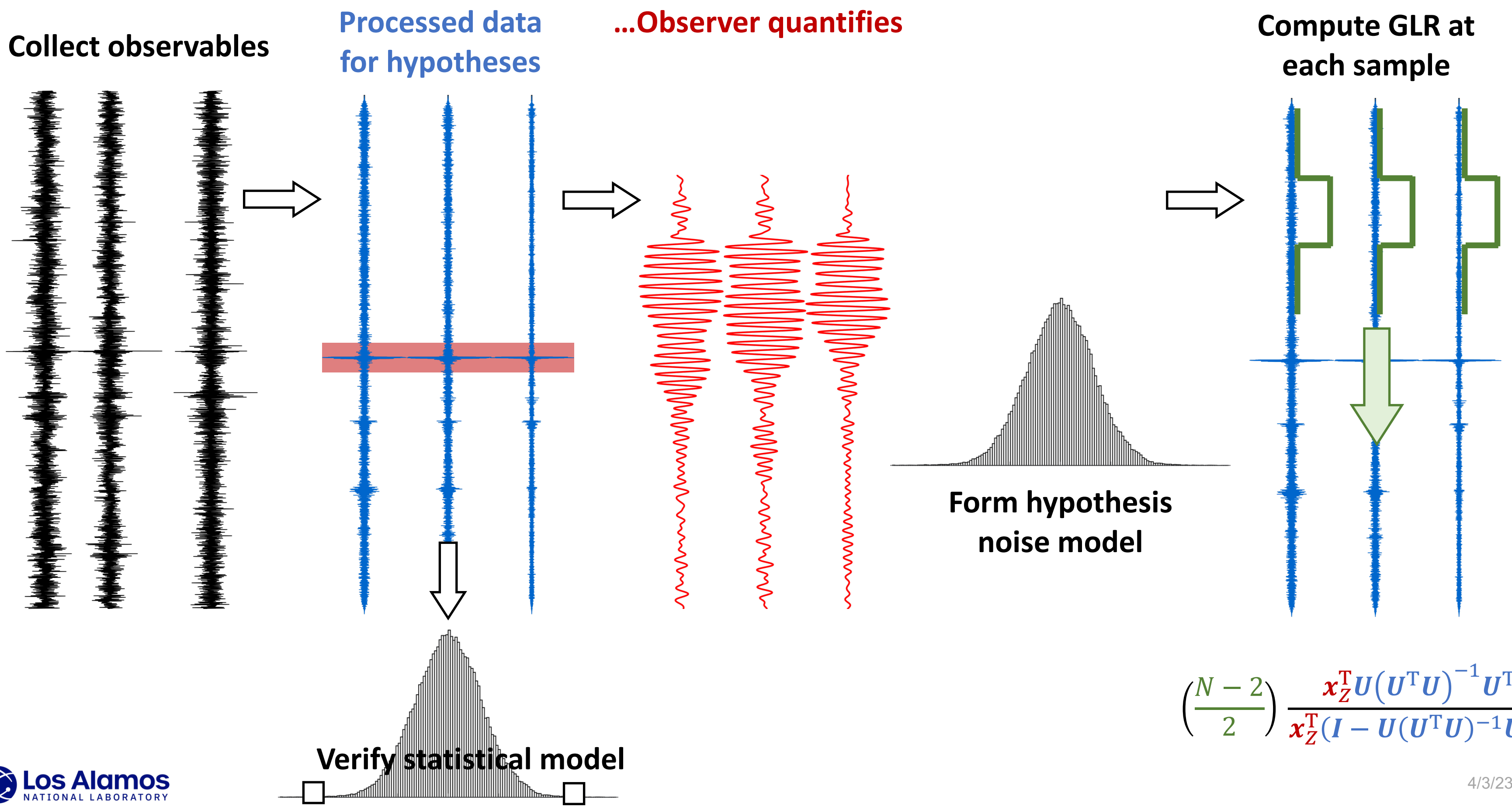
1. **Scope**: Elements of detection theory, with focus on the practical and computational aspects of digital waveform detection.
2. **Non-Scope**: You will not watch me debug code in real time; no one wants to see that.
3. **Approach**: start to finish demonstration of a digital Rayleigh wave detector:
    1. The retrograde, elliptically polarized motion (REPM) signal detector, with *possible additional reference to*:
    2. Computation of p-values
    3. Power detectors (STA/LTA, F-detectors)
    4. Correlation detectors
    5. Eigenmotion detectors
4. **A Hypothetical Course Catalog Description**: common tests for detectors; efficient, point-wise matrix inversions (2x2, 3x3); fitting densities to normalized histograms; setting false alarm rate thresholds; Bayesian estimate of parameters and thresholds; ~~physical interpretation of noncentrally parameters~~; ~~empirical quantification of detector performance~~ ← [Requires waveform injection tutorial]

Bottom Line Up Front (High-Fidelity BLUF): We Will (1/4)



$$\left(\frac{N-2}{2}\right)\frac{\boldsymbol{x}_Z^{\mathrm{T}}\boldsymbol{U}\left(\boldsymbol{U}^{\mathrm{T}}\boldsymbol{U}\right)^{-1}\boldsymbol{U}^{\mathrm{T}}\boldsymbol{x}_Z}{\boldsymbol{x}_Z^{\mathrm{T}}(\boldsymbol{I}-\boldsymbol{U}(\boldsymbol{U}^{\mathrm{T}}\boldsymbol{U})^{-1}\boldsymbol{U}^{\mathrm{T}})\boldsymbol{x}_Z}$$

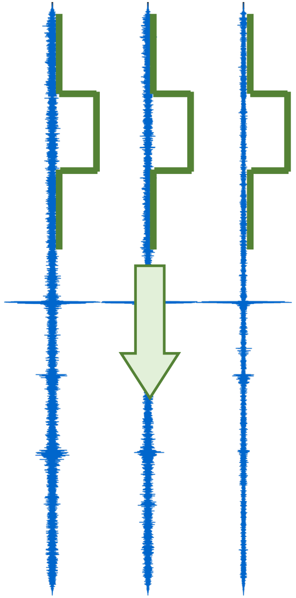
Bottom Line Up Front (High-Fidelity BLUF): We Will (2/4)



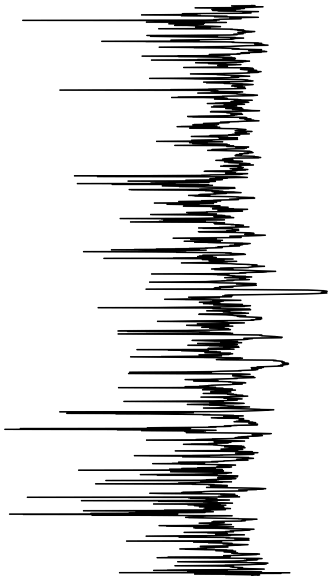
$$\left(\frac{N-2}{2}\right)\frac{\boldsymbol{x}_Z^{\mathrm{T}}\boldsymbol{U}\left(\boldsymbol{U}^{\mathrm{T}}\boldsymbol{U}\right)^{-1}\boldsymbol{U}^{\mathrm{T}}\boldsymbol{x}_Z}{\boldsymbol{x}_Z^{\mathrm{T}}(\boldsymbol{I}-\boldsymbol{U}(\boldsymbol{U}^{\mathrm{T}}\boldsymbol{U})^{-1}\boldsymbol{U}^{\mathrm{T}})\boldsymbol{x}_Z}$$

Bottom Line Up Front (High-Fidelity BLUF): We Will (3/4)

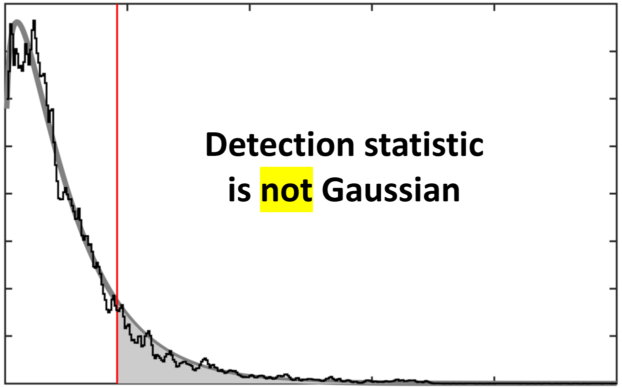
Compute GLR at each sample



Output detection statistic

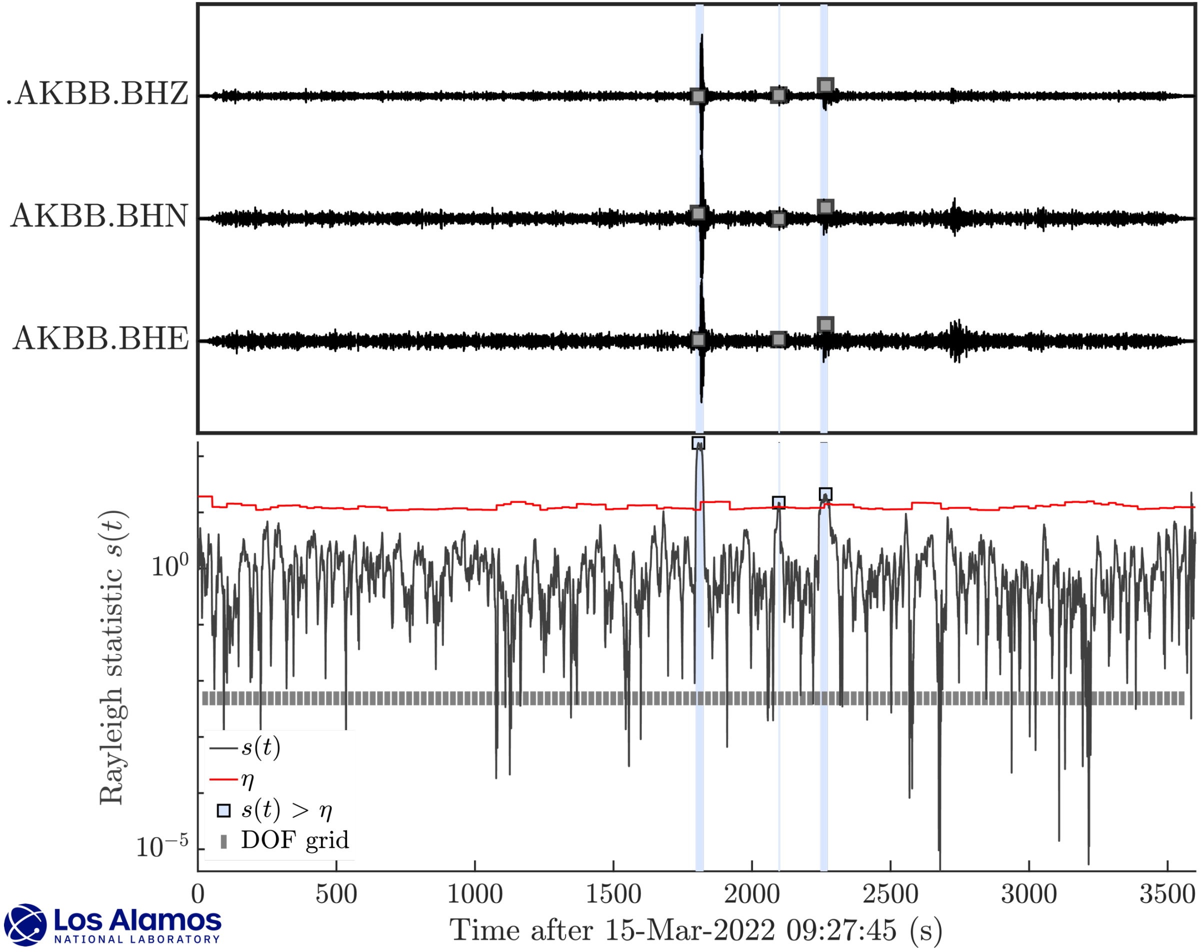


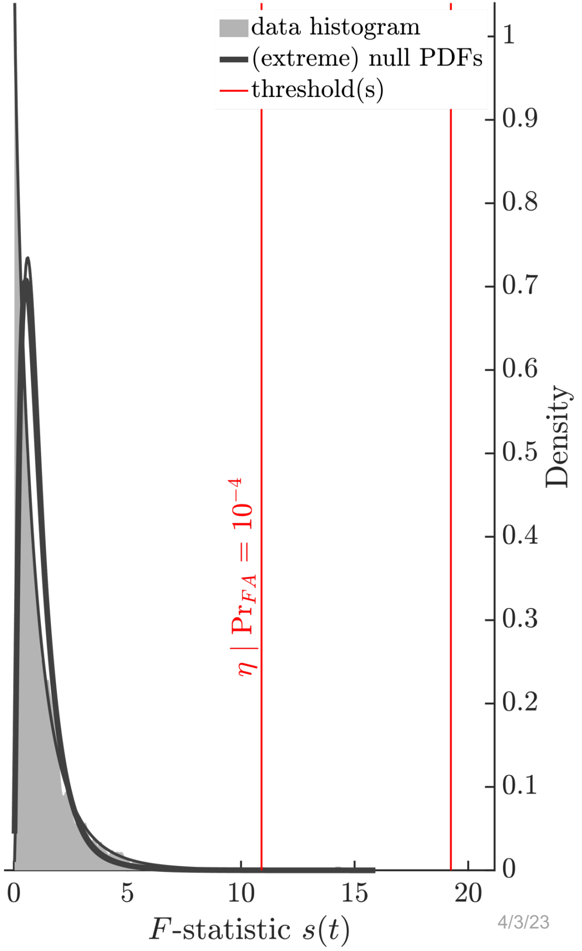
Bin detection statistic, estimate parameters, compute threshold for CFAR



CFAR is not a p-value

Bottom Line Up Front (High-Fidelity BLUF): We Will (4/4)





Los Alamos NATIONAL LABORATORY

Module 0: Meta-Detection

Detectors are Binary Hypothesis Tests that Quantify the Scientific Method

The Scientific Method[1]

1. Define a question / observation
2. Form a prediction (a hypothesis)
3. Gather data
4. Analyze the Data
5. Accept or refute your prediction (hypothesis)

Signal Detection[2]

1. Does the region of interest host a target source?
2. Form competing hypotheses: $\mathcal{H}_0$: sensors record noise + background; $\mathcal{H}_1$: sensors *also* record target sources.
3. Input digital data into competing models for each hypothesis.
4. Collect data and form a generalized likelihood ratio (GLRT).
5. Compare detection statistics to thresholds to declare which hypothesis is true.

[1]American Museum of Natural History

[2]From binary hypothesis tests

The Five Steps of Detection (think Scientific Method)

Single modality detection: terminology, concepts

$$\mathcal{H}_0\colon\ \boldsymbol{x} = \text{noise} + \text{background}$$

$$\mathcal{H}_1\colon\ \boldsymbol{x} = \text{noise} + \text{background} + \text{signal}$$

$$\frac{\widehat{\mathrm{PDF}}(\boldsymbol{x}|\mathcal{H}_1)}{\widehat{\mathrm{PDF}}(\boldsymbol{x}|\mathcal{H}_0)} = s(\boldsymbol{x}) \underset{\mathcal{H}_0}{\overset{\mathcal{H}_1}{\gtrless}} t \qquad \mathcal{H}_1 \rightarrow u = 1, \quad \mathcal{H}_0 \rightarrow u = 0$$

Source | **Data Models** | **GLR** | **Detector** | **Decision**

Origin of observables

Form competing hypotheses

Gather data / observables

Analyze data/observables

Accept or reject initial hypothesis

Density processing: estimate distributional parameters that shape density and histograms

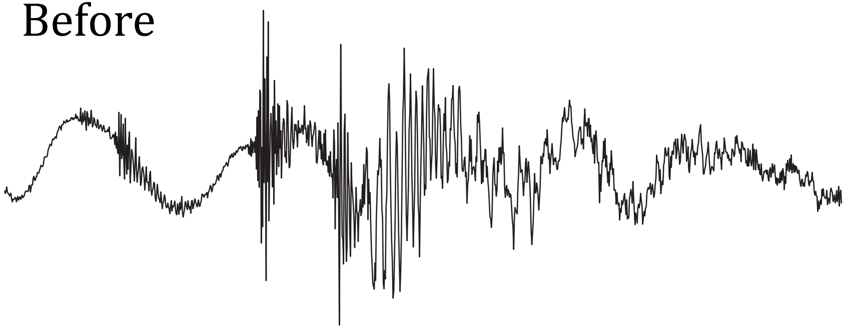
Data pre-processing: detrend, taper and bandpass filter (possibly in a filter bank)

Module 1:
Before Signal Detection

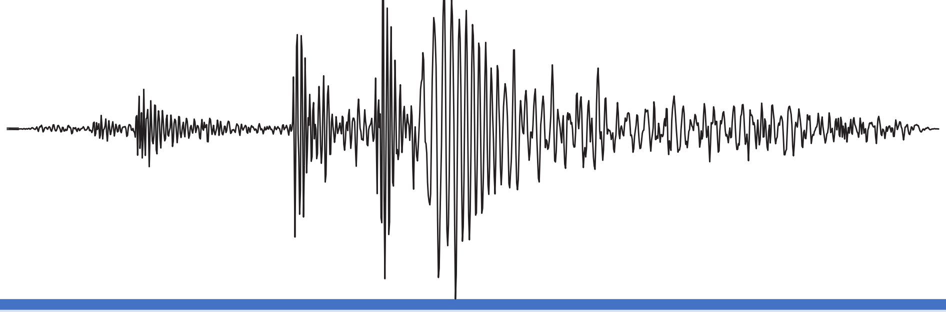
Module 1: Before Signal Detection (1/7)

- For each channel of waveform data:
  - Select time segments such that noise statistics remain static (suggest 15 min to $\leq$ 2 hr)
  - Decimate or **resample** (if required) *before* you:
  - **Detrend the data** to remove trend line or mean
    - *Possibly* high-pass filter data to remove very long period trends
  - **Taper the data** ends to prevent spectral leakage
  - **Process data with filter(s)** over sensical bands (e.g., $\leq$ 85% Nyquist).

Before



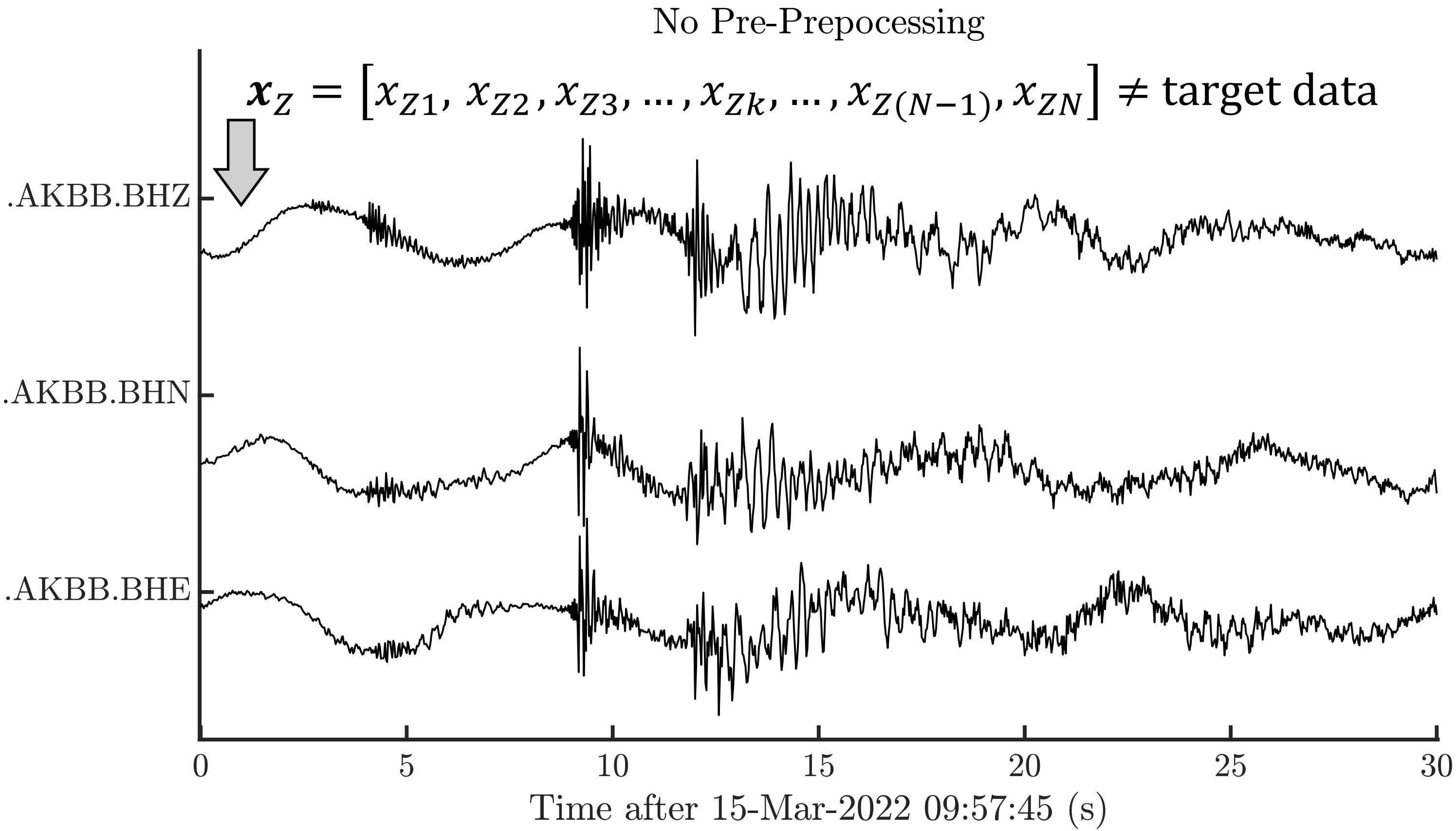
After (ready for signal detection)



Perform signal detection against **processed** data

Module 1: Before Signal Detection (2/7)

Lecture
lanlTracePlot.m

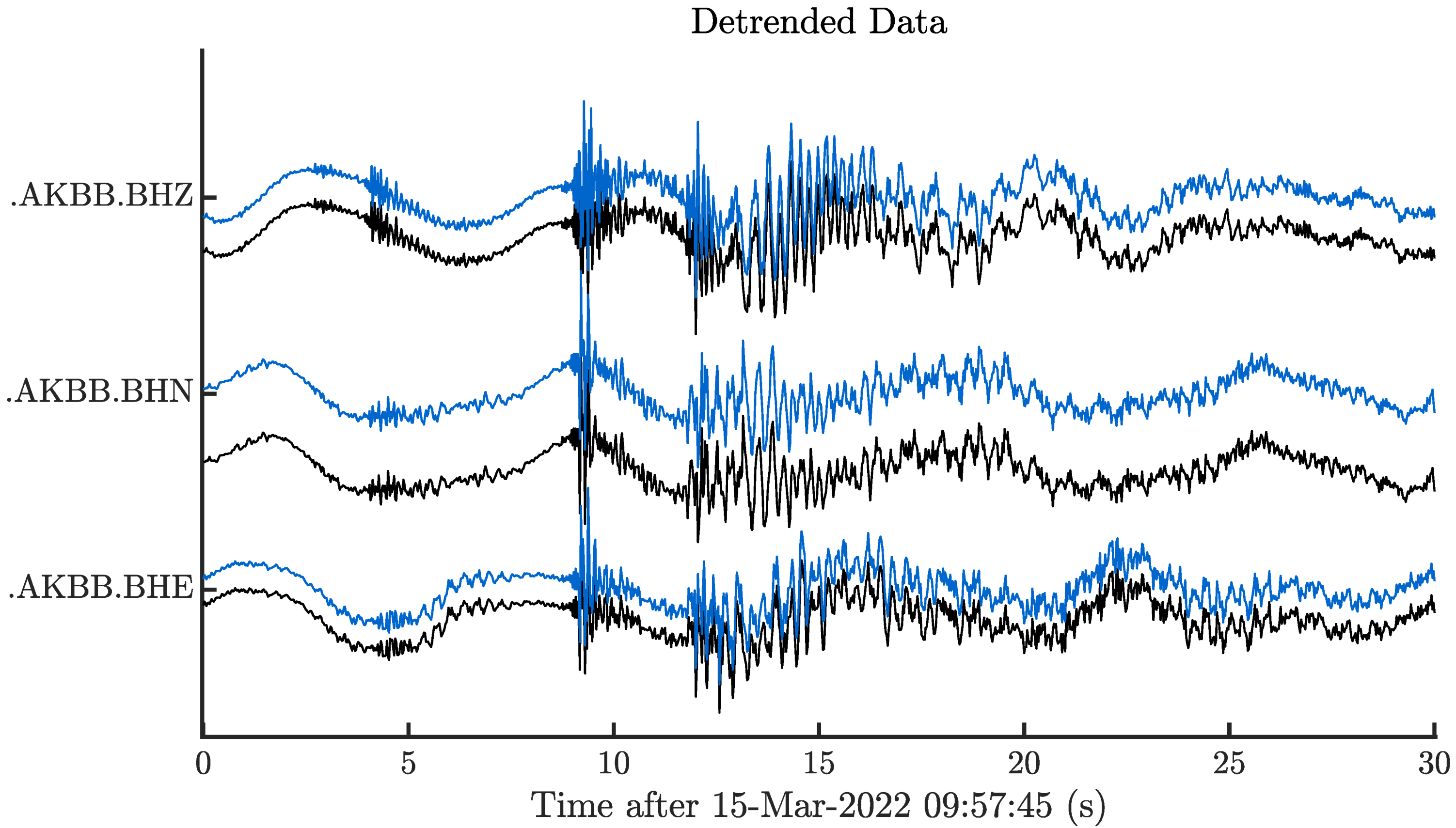


Do **not** usually perform signal detection against *raw* data

Module 1: Before Signal Detection (3/7)

Lecture

lanlTraceDetrend.m

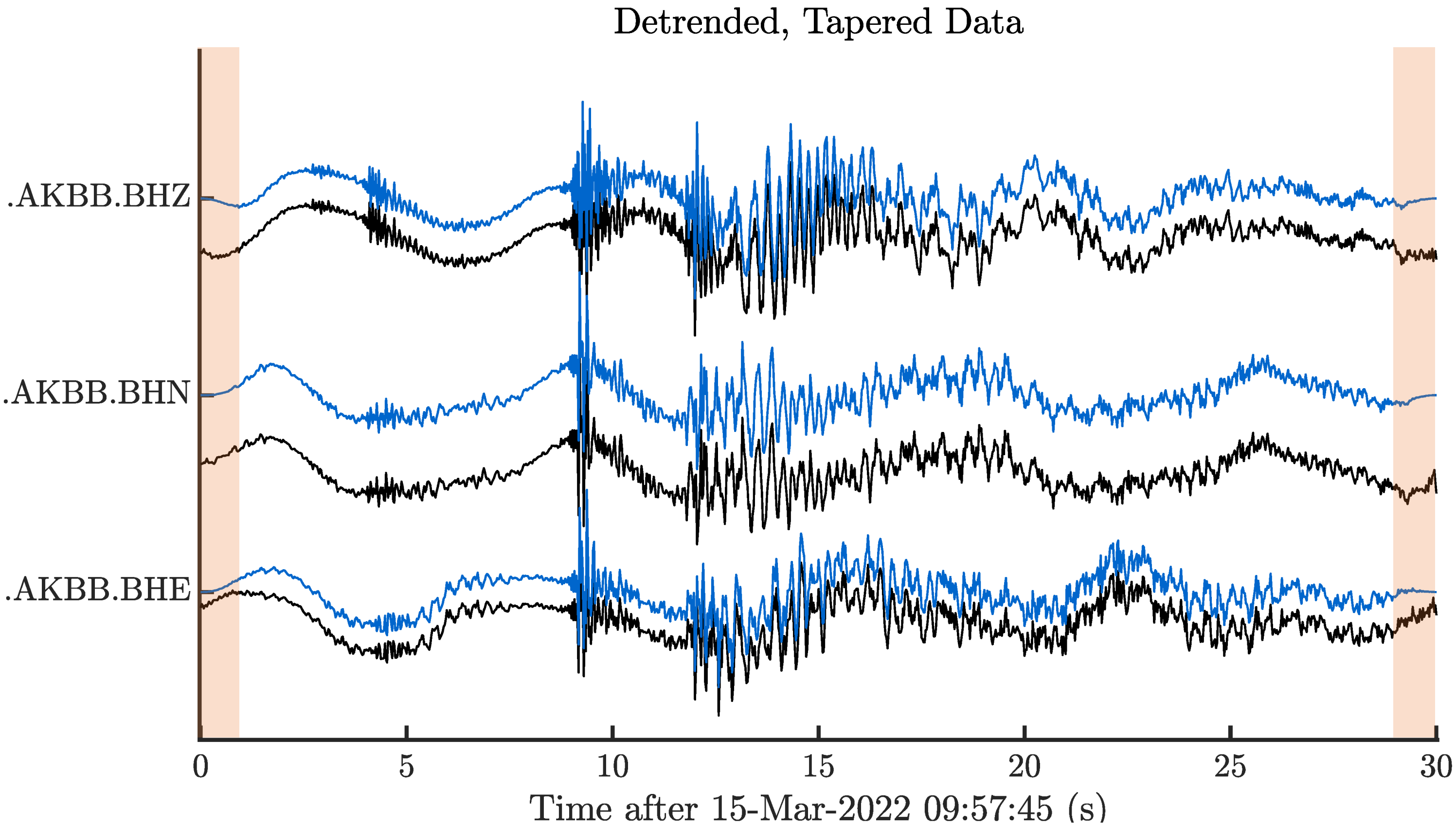


Detrend target data to remove trends and mean prior to tapering

Module 1: Before Signal Detection (4/7)

Lecture

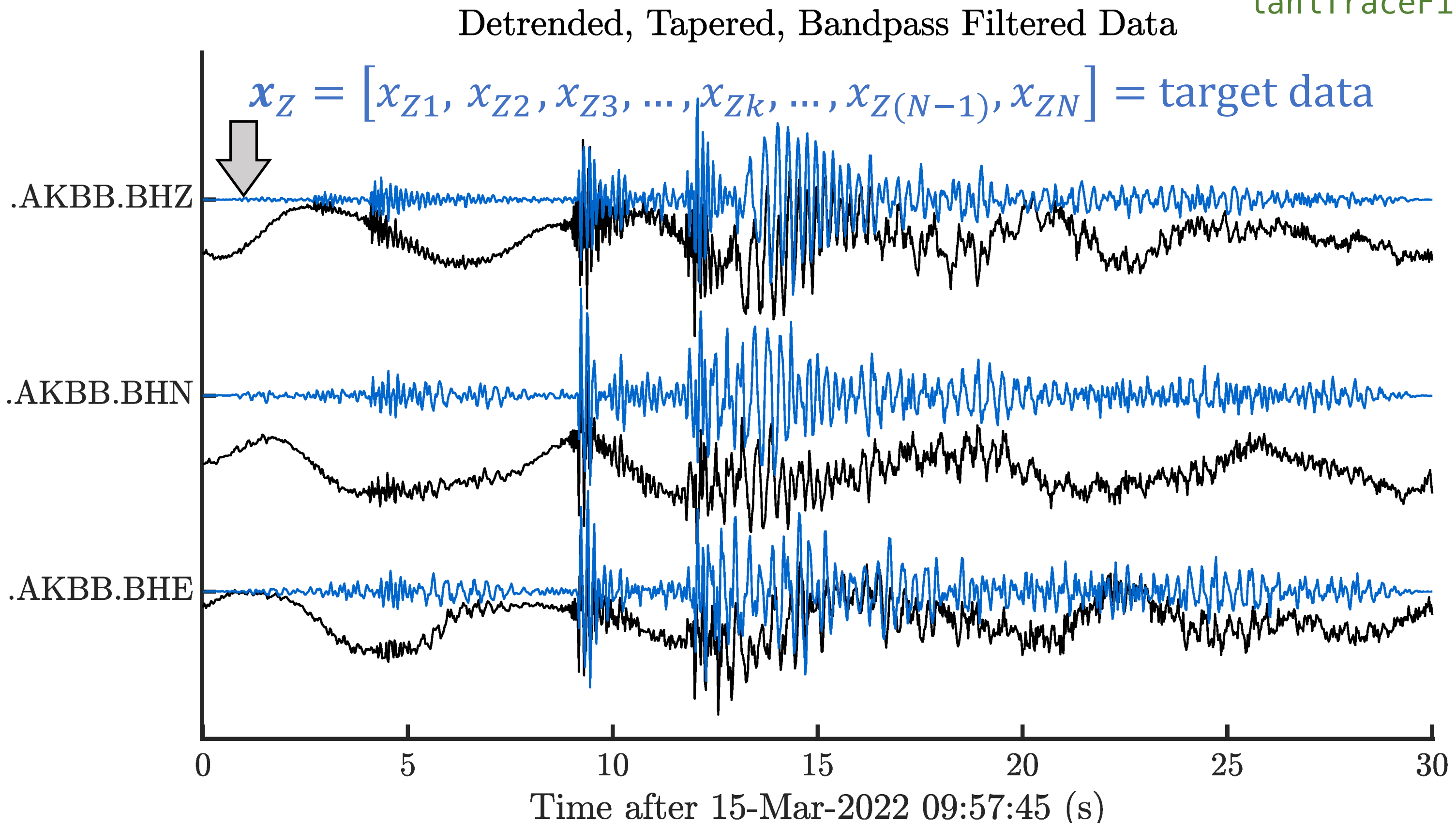
lanlTukeyTaper.m, lanlTraceDetrend.m



Taper target data at the ends to prevent spectral leakage upon filtering

Module 1: Before Signal Detection (5/7)

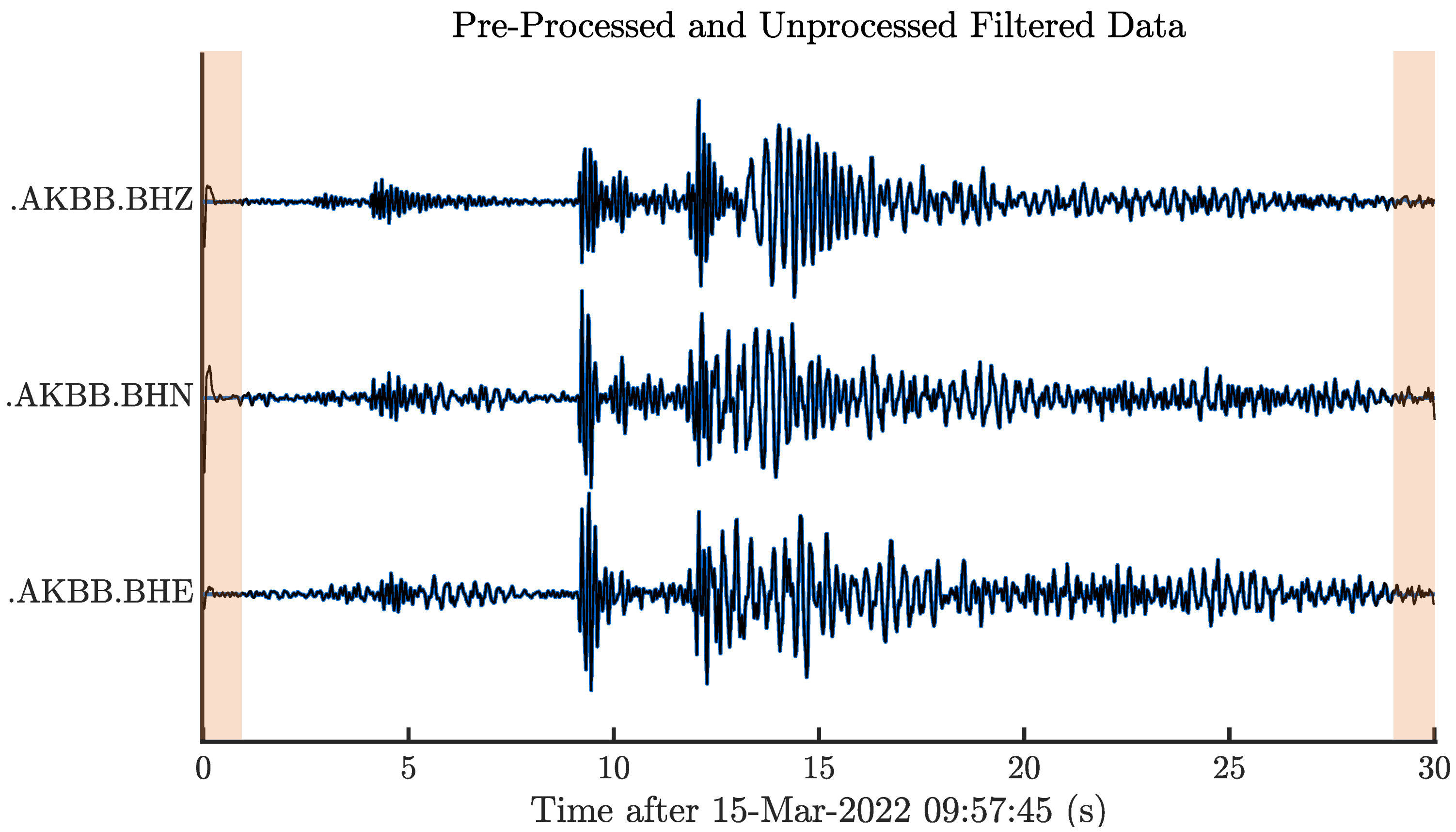
`lanlTukeyTaper.m`, `lanlTraceDetrend.m`, `lanlTraceFiltButter.m`



Process data with **minimum phase bandpass filter** to minimize acausal time shifts

Module 1: Before Signal Detection (6/7)

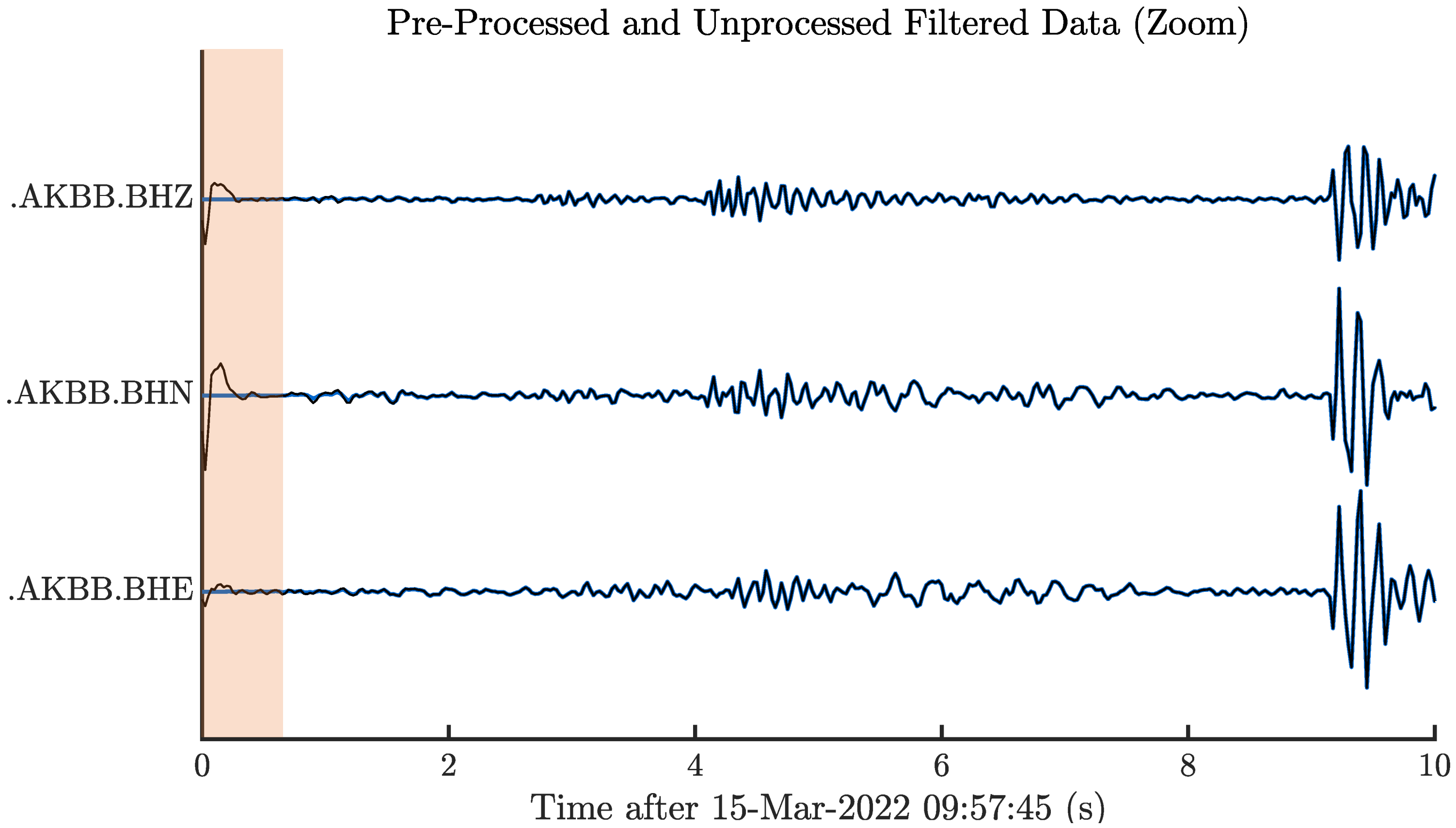
lanlTracePlot.m



Compare correctly processed data with incorrectly pre-processed data

Module 1: Before Signal Detection (7/7)

lanlTracePlot.m

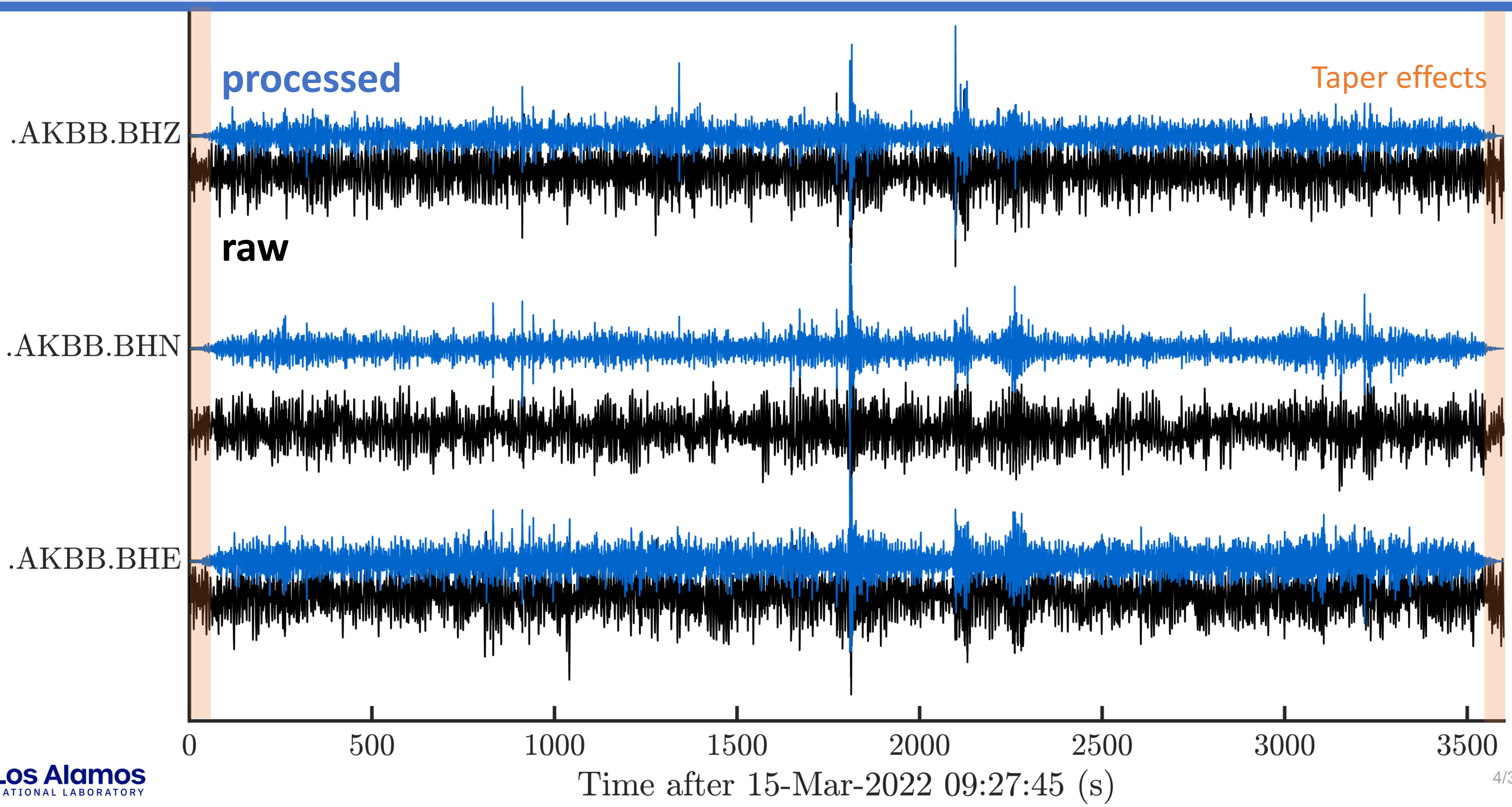


A signal detector could erroneously declare that unprocessed end-samples include signal

Module 2:
Verify Statistical Assumptions

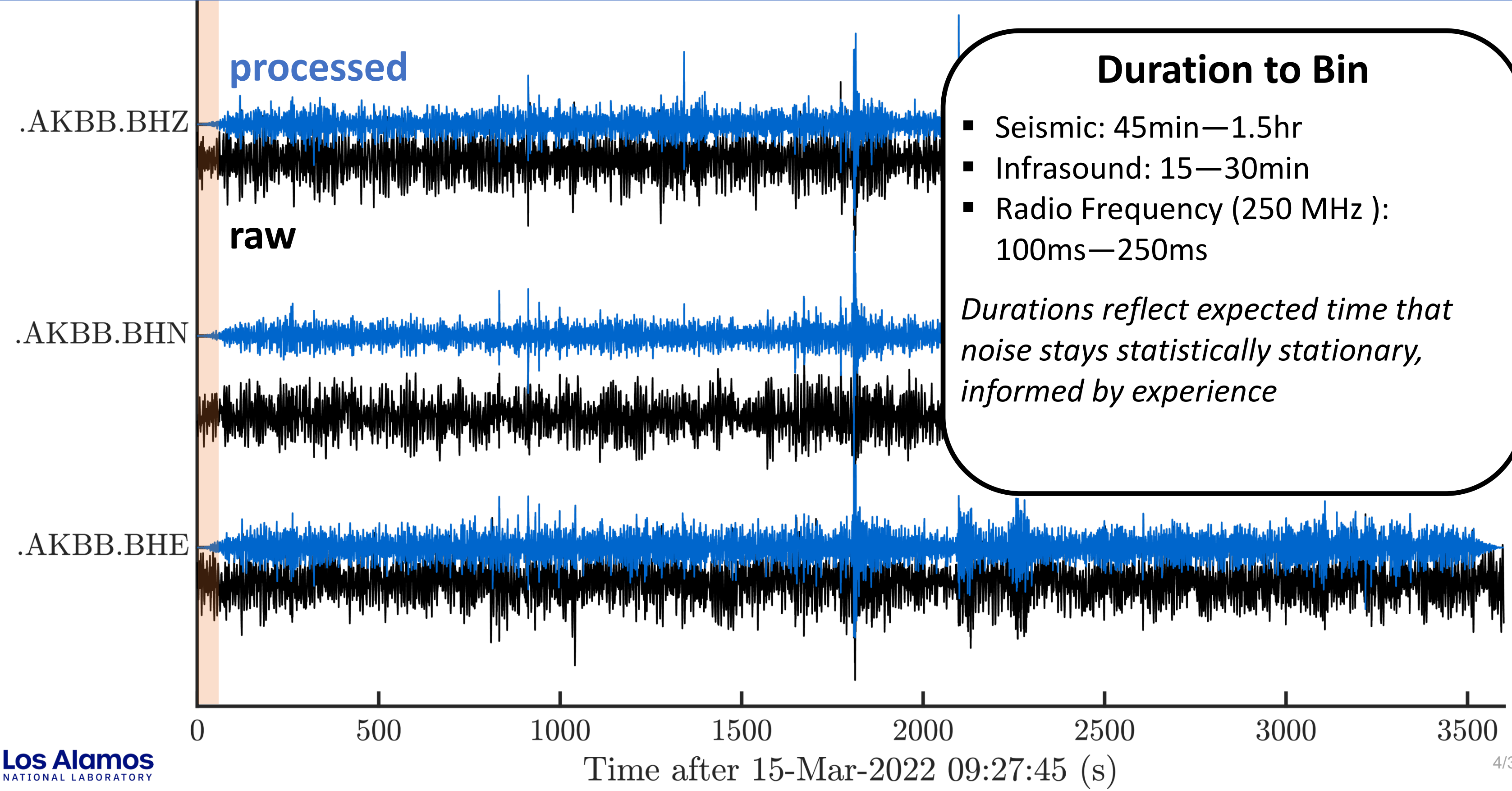
Module 2: Verify Statistical Assumptions (1/8)

Test **normality** of ~1 hr of the target data



Module 2: Verify Statistical Assumptions (2/8)

Verify **duration** of target data to test is within “best-practice” durations



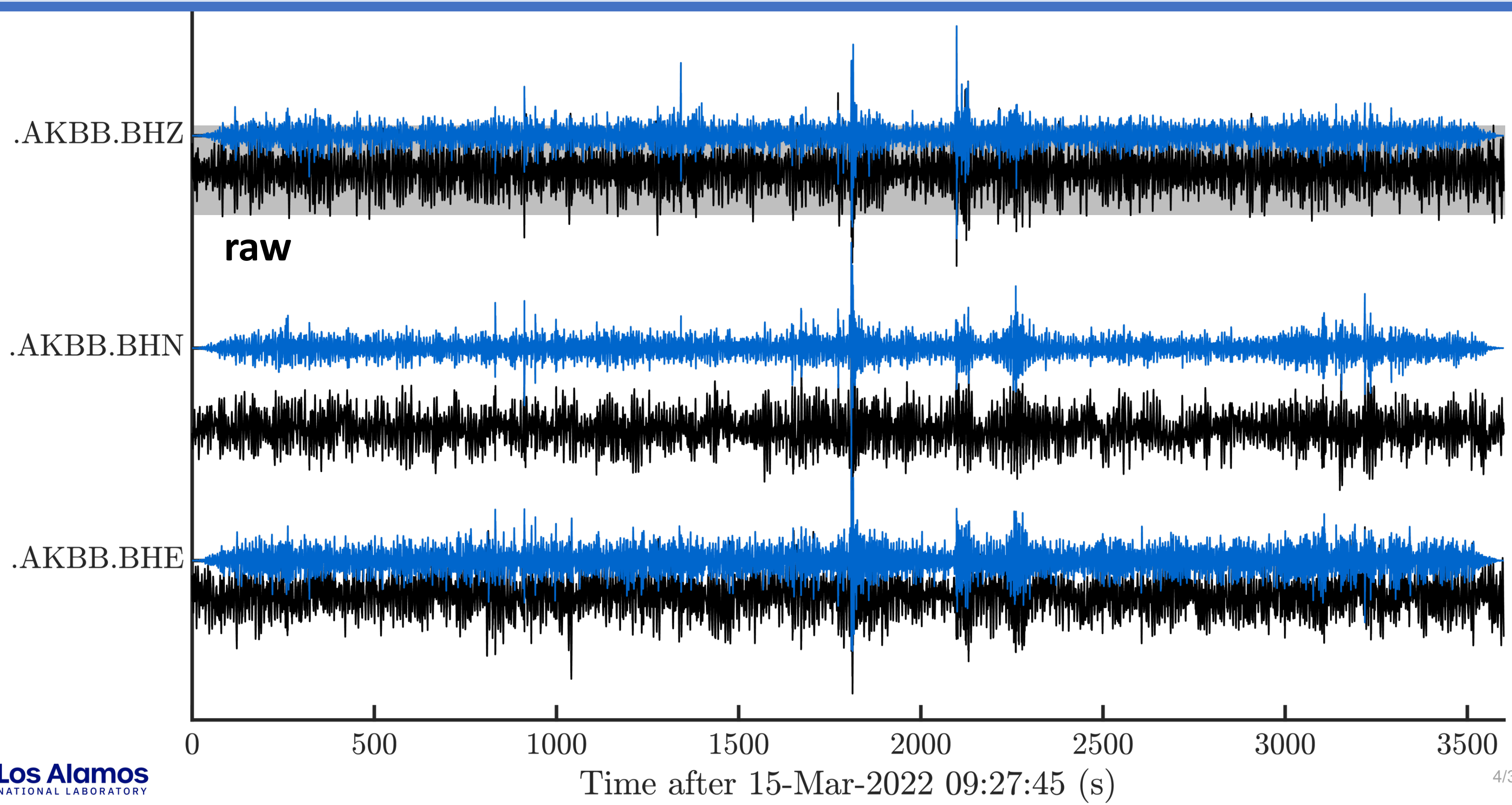
Duration to Bin

- Seismic: 45min—1.5hr
- Infrasound: 15—30min
- Radio Frequency (250 MHz): 100ms—250ms

Durations reflect expected time that noise stays statistically stationary, informed by experience

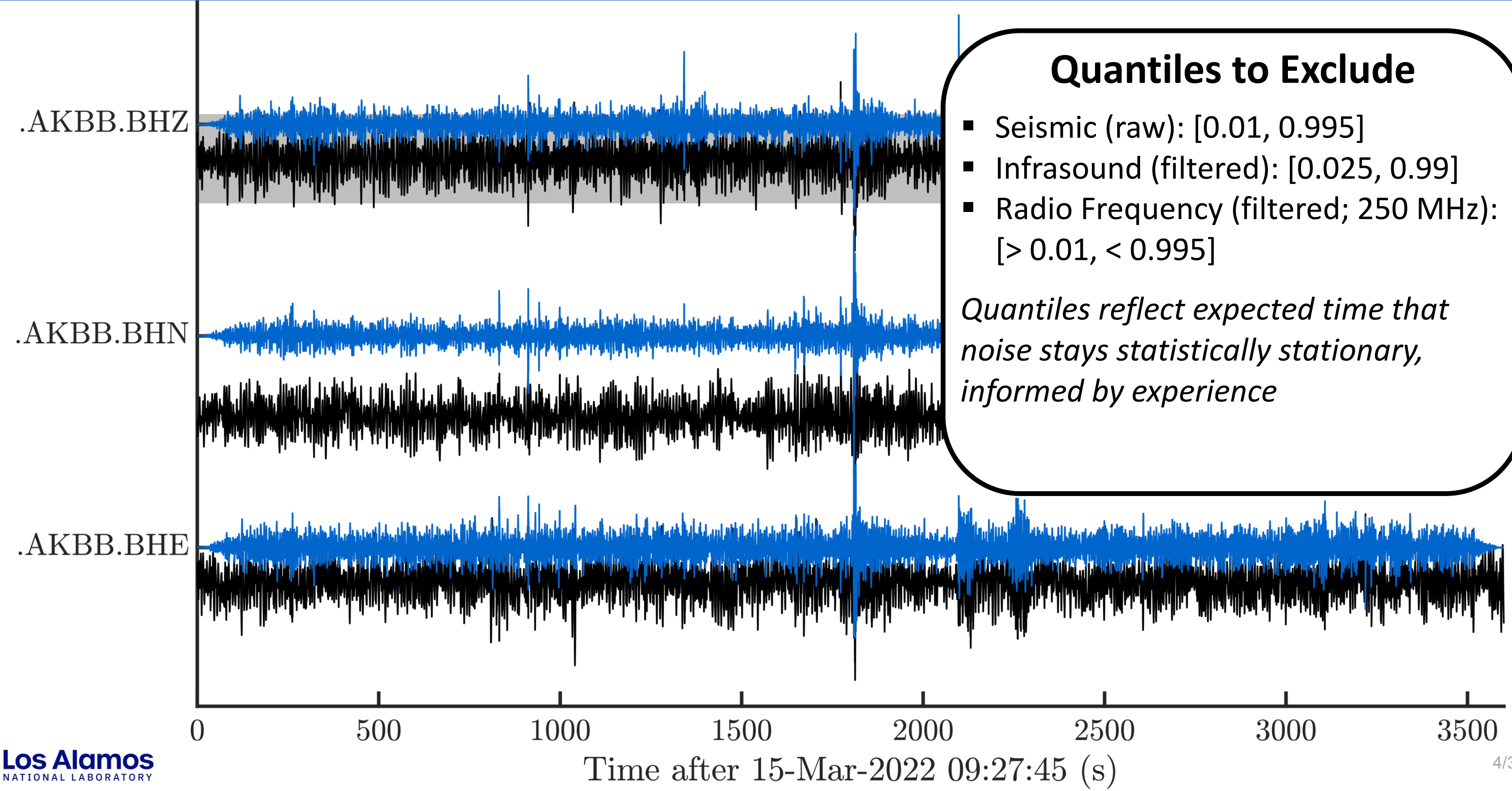
Module 2: Verify Statistical Assumptions (3/8)

Do not bin all data; **exclude** small and large **quantiles** [0.01, 0.995]



Module 2: Verify Statistical Assumptions (4/8)

Verify that **quantiles** approximately meet "best practices" to **exclude**



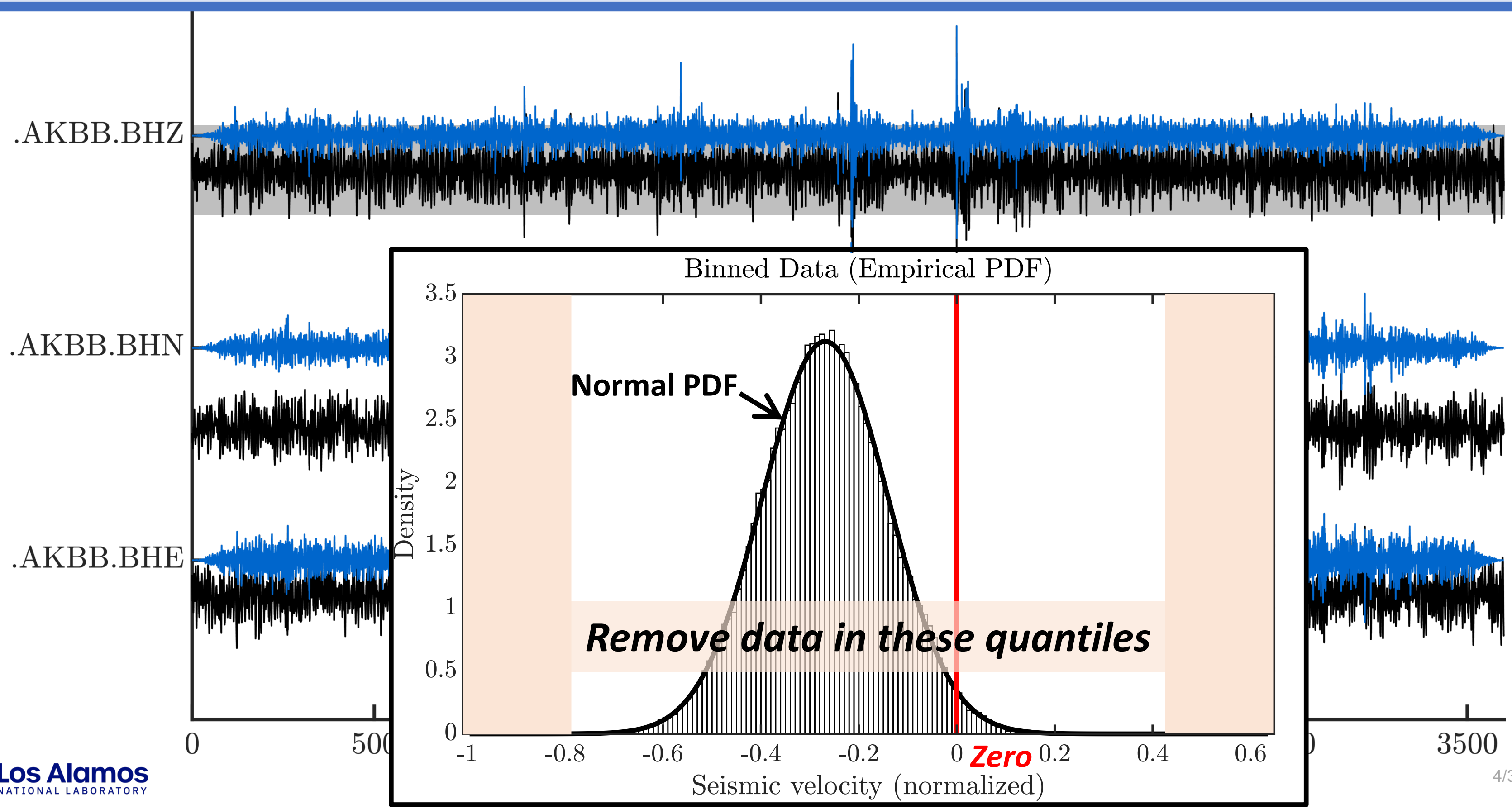
Quantiles to Exclude

- Seismic (raw): [0.01, 0.995]
- Infrasound (filtered): [0.025, 0.99]
- Radio Frequency (filtered; 250 MHz): [> 0.01, < 0.995]

Quantiles reflect expected time that noise stays statistically stationary, informed by experience

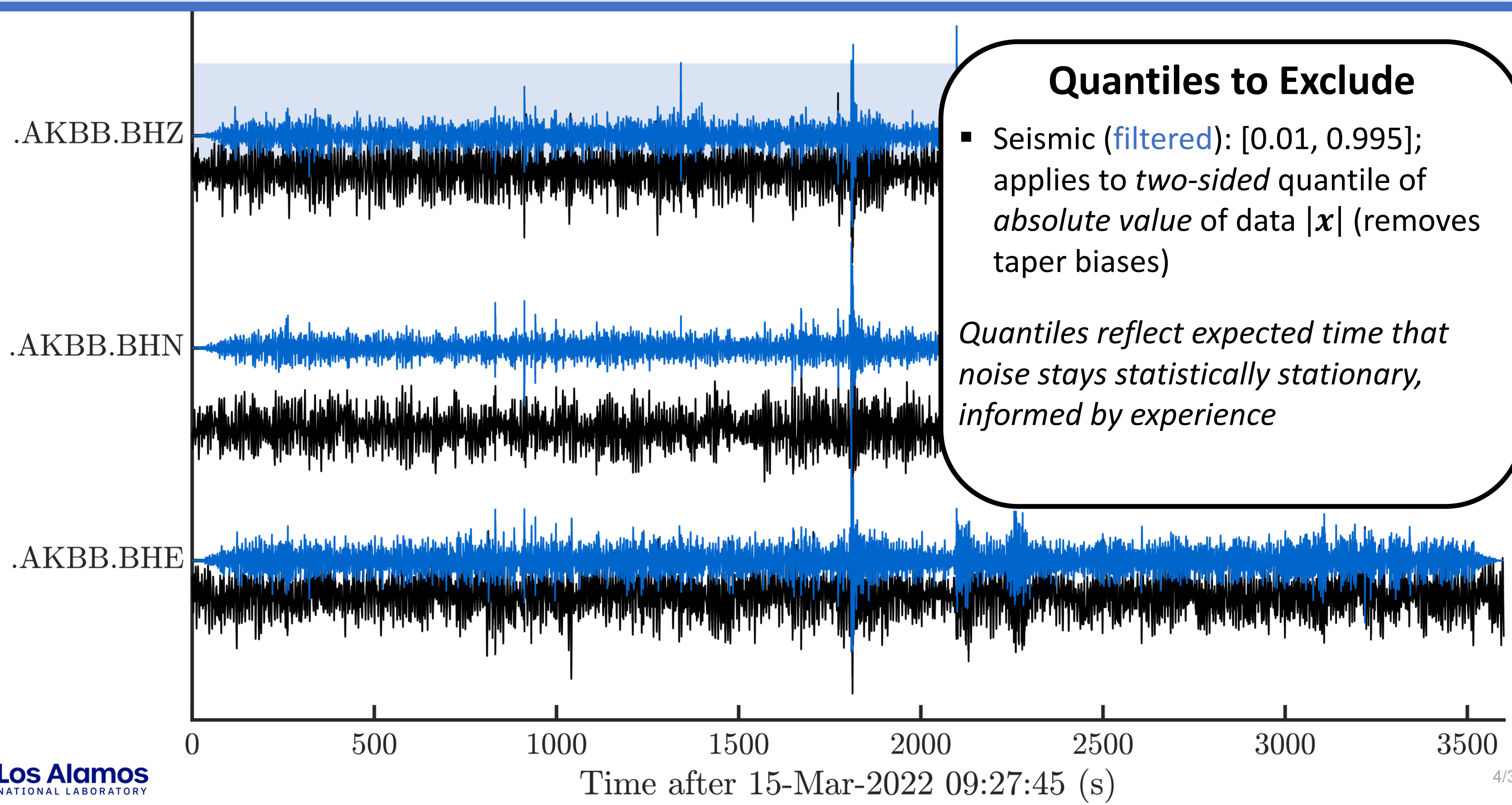
Module 2: Verify Statistical Assumptions (5/8)

Bin; raw noise data appears **normal/Gaussian** but shows high variance and a **non-zero mean**



Module 2: Verify Statistical Assumptions (6/8)

Longer duration data, contains **same** explosion-sourced event



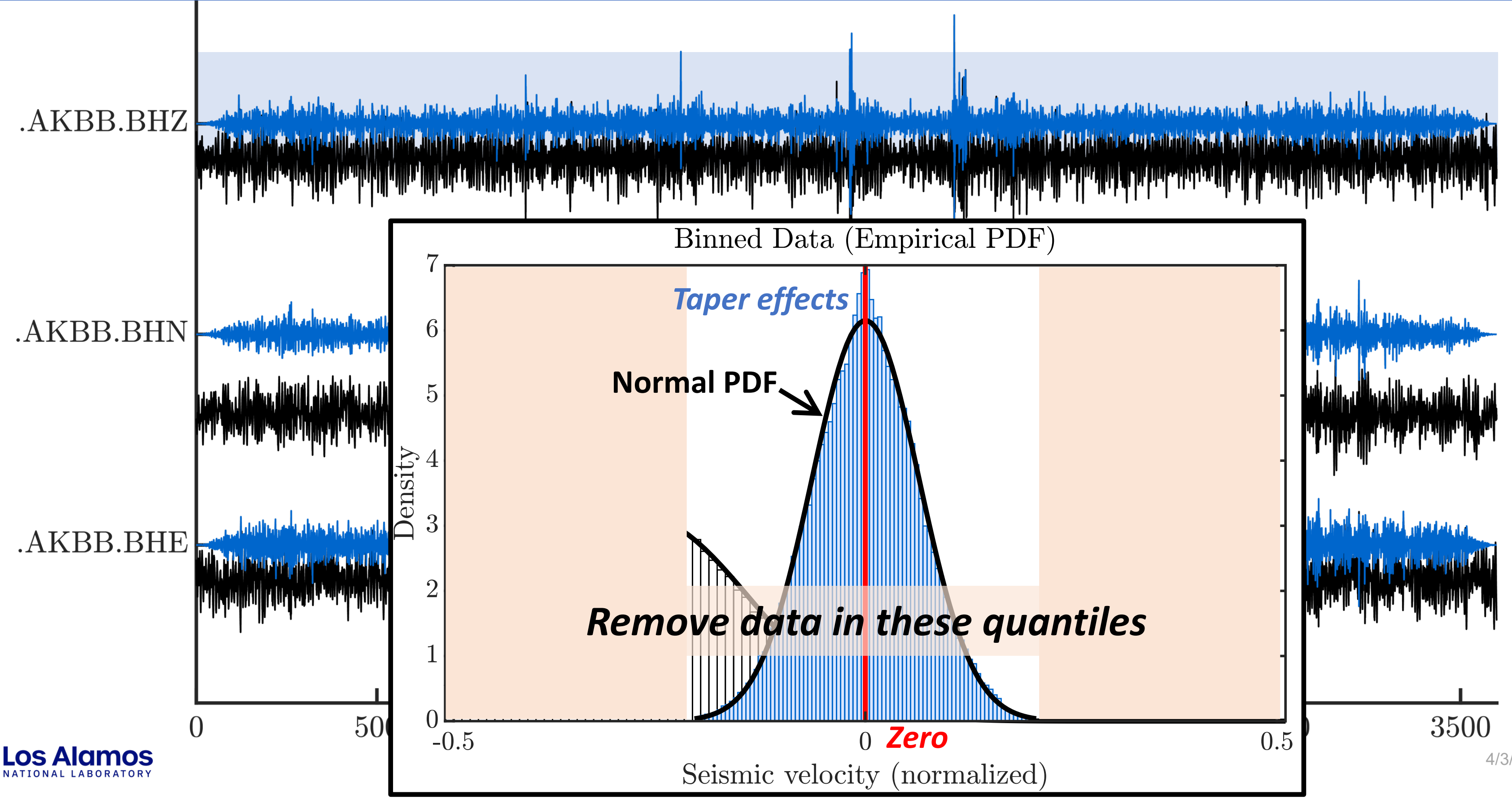
Quantiles to Exclude

- Seismic (filtered): [0.01, 0.995]; applies to *two-sided* quantile of *absolute value* of data $|\boldsymbol{x}|$ (removes taper biases)

Quantiles reflect expected time that noise stays statistically stationary, informed by experience

Module 2: Verify Statistical Assumptions (7/8)

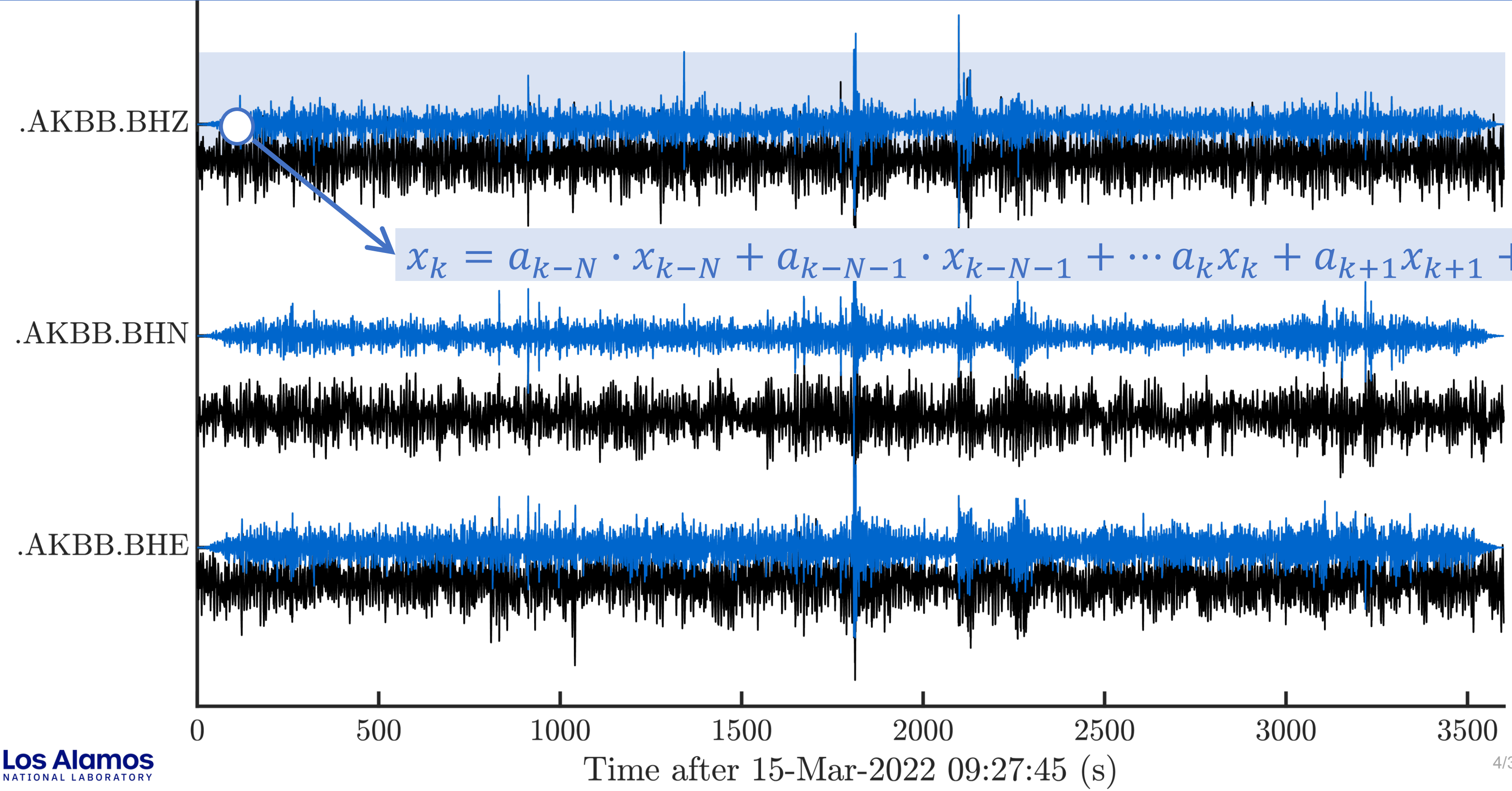
Bin; processed data is **normal/Gaussian** but shows lower variance and a **zero mean**



Los Alamos NATIONAL LABORATORY

Module 2: Verify Statistical Assumptions (8/8)

Caveat: Filtering induces *correlation*. Each sample is a weighted combination of its neighbors.



$$x_k = a_{k-N} \cdot x_{k-N} + a_{k-N-1} \cdot x_{k-N-1} + \cdots a_k x_k + a_{k+1} x_{k+1} + \cdots$$

Module 3:
Form Competing Data Hypothesis

Status Update: What we've Done Already

Single modality detection: terminology, concepts

$\mathcal{H}_0$: $\boldsymbol{x} =$ noise + background

$\mathcal{H}_1$: $\boldsymbol{x} =$ noise + background + signal

$$\frac{\widehat{\mathrm{PDF}}(\boldsymbol{x}|\mathcal{H}_1)}{\widehat{\mathrm{PDF}}(\boldsymbol{x}|\mathcal{H}_0)} = s(\boldsymbol{x}) \underset{\mathcal{H}_0}{\overset{\mathcal{H}_1}{\gtrless}} t$$

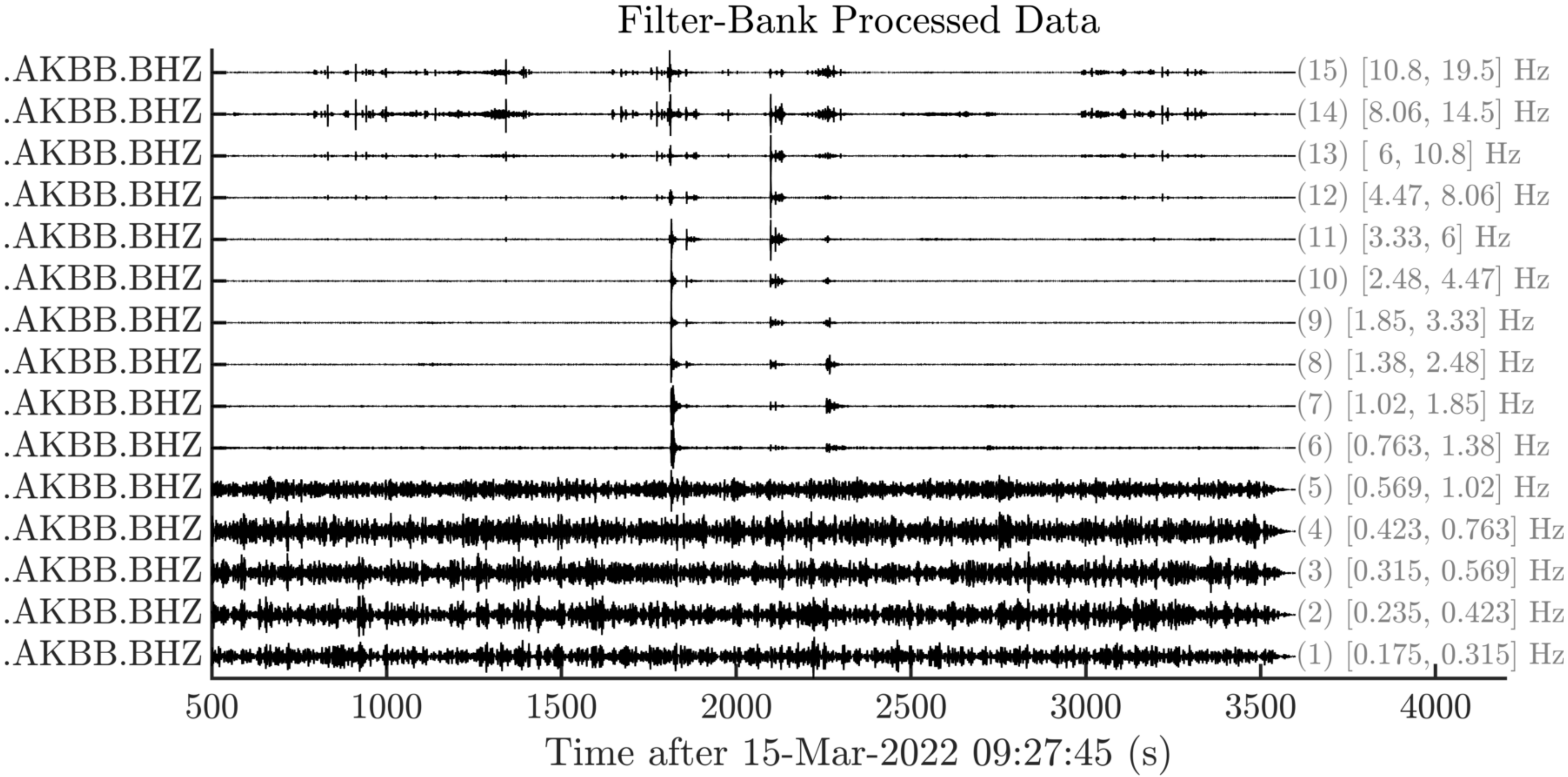
$\mathcal{H}_1 \rightarrow u = 1$

$\mathcal{H}_0 \rightarrow u = 0$

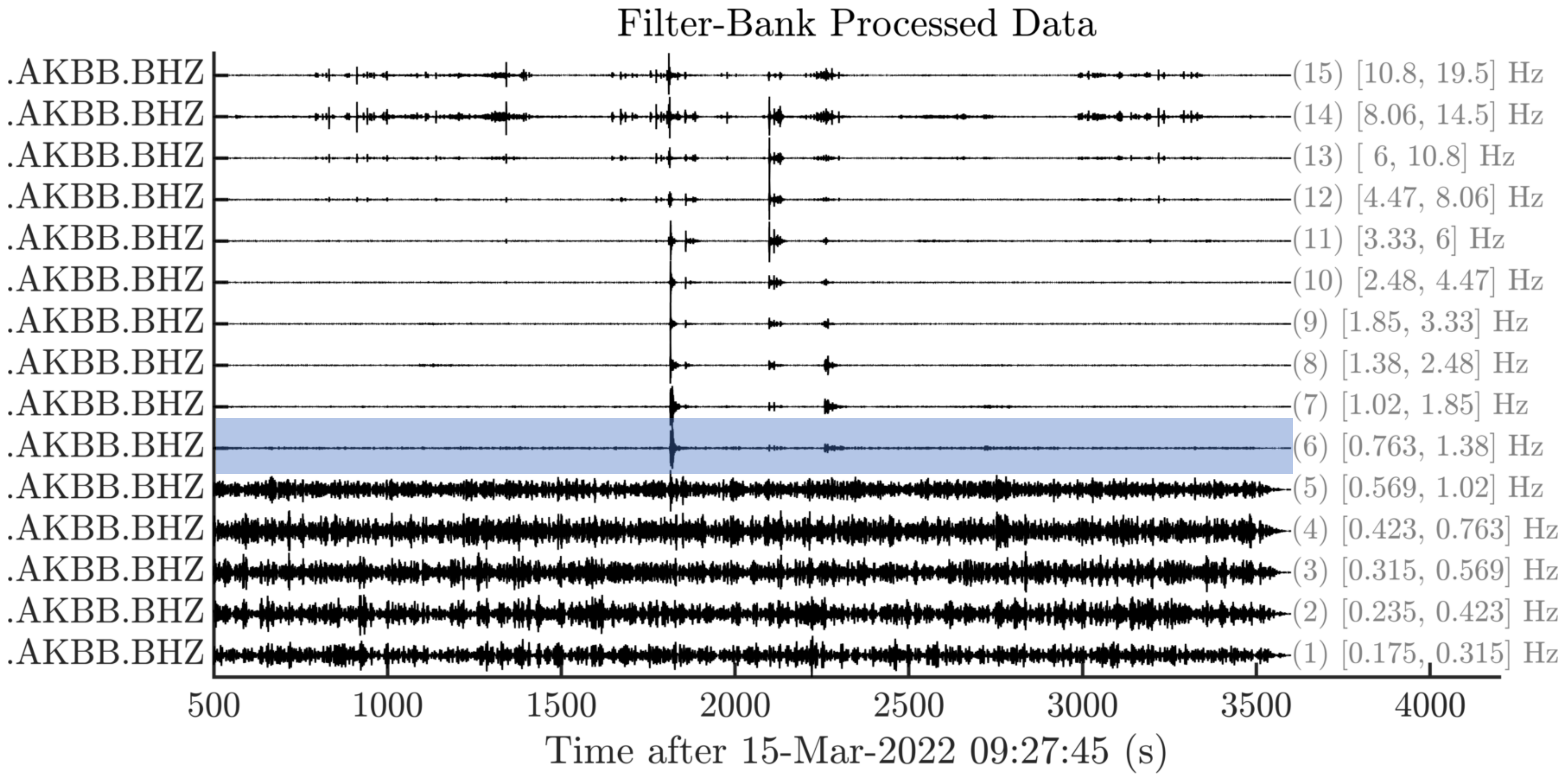
Source | **Data Models** | **GLR** | **Detector** | **Decision**

Data pre-processing: detrend, taper and bandpass filter (possibly in a filter bank)

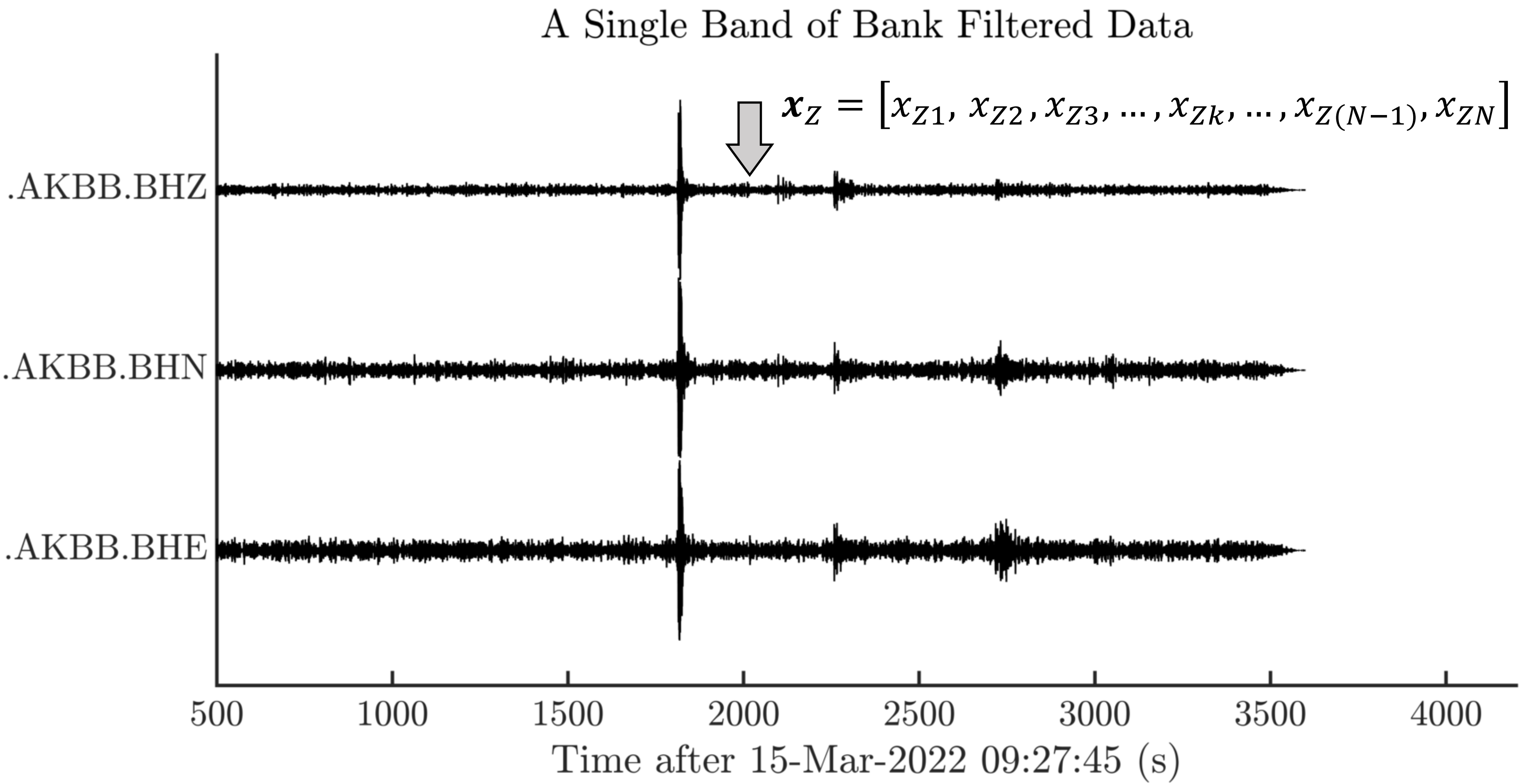
Module 3: Form Competing Data Hypothesis (1/21)



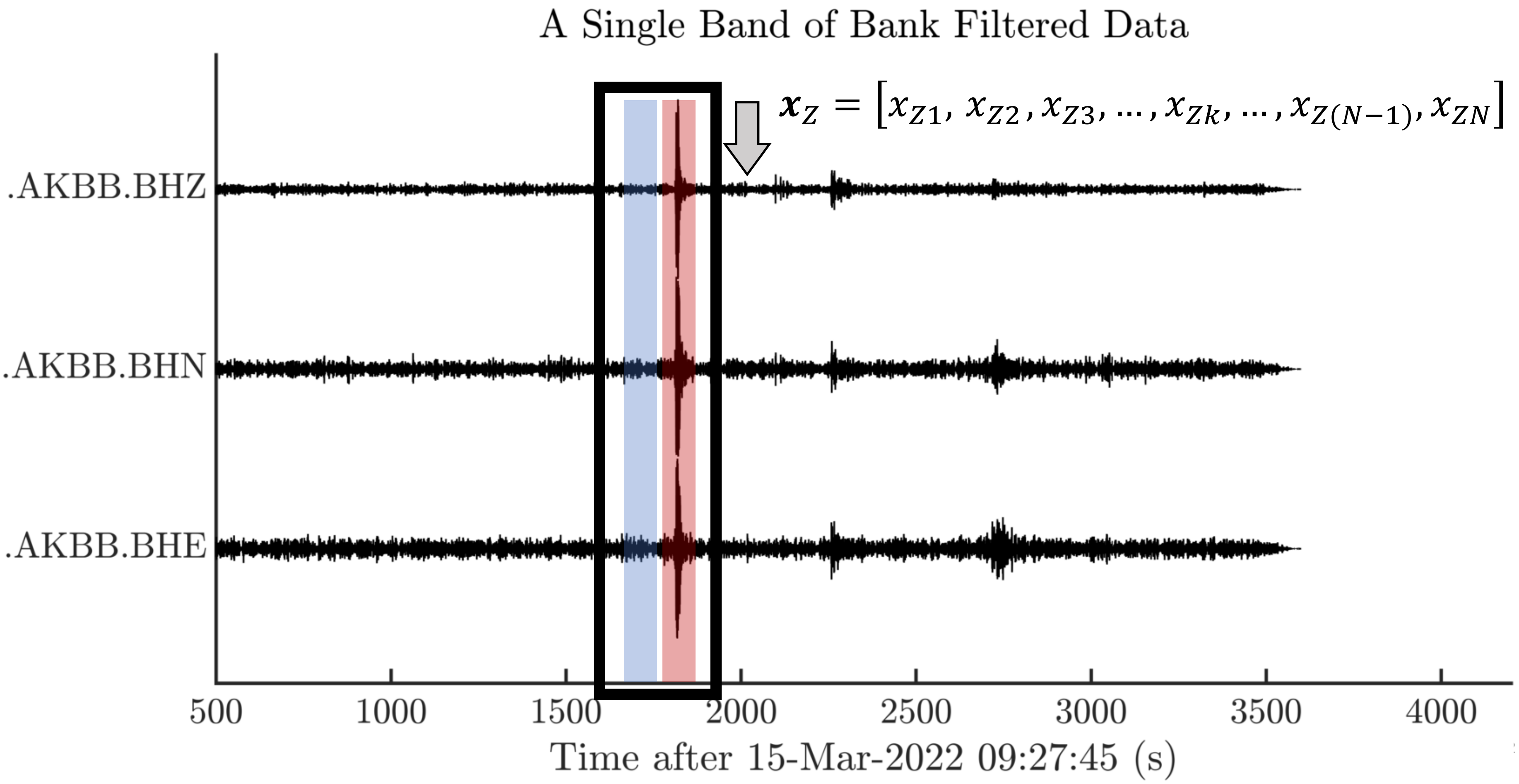
Module 3: Form Competing Data Hypothesis (2/21)



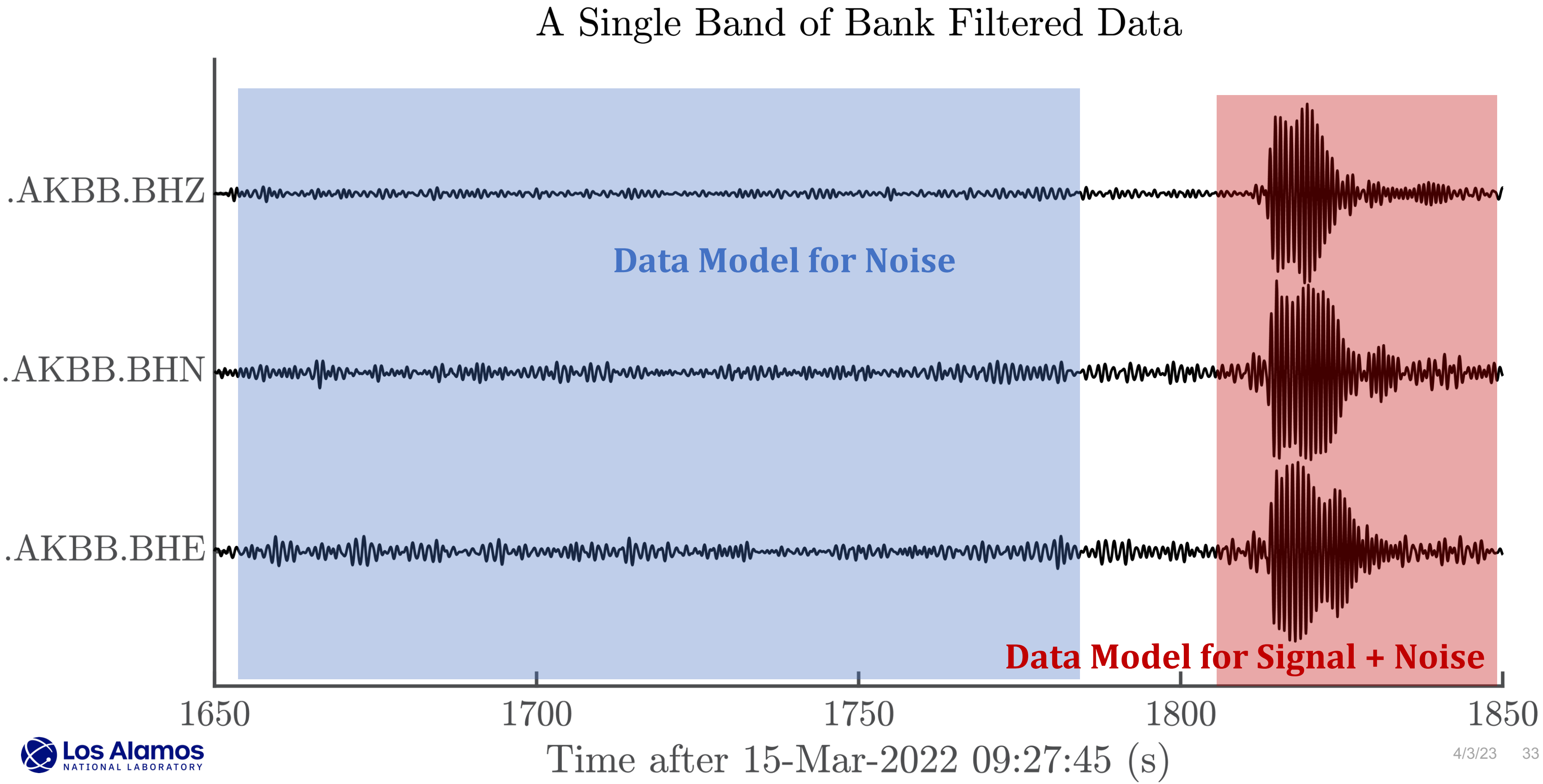
Module 3: Form Competing Data Hypothesis (2/18)



Module 3: Form Competing Data Hypothesis (3/18)



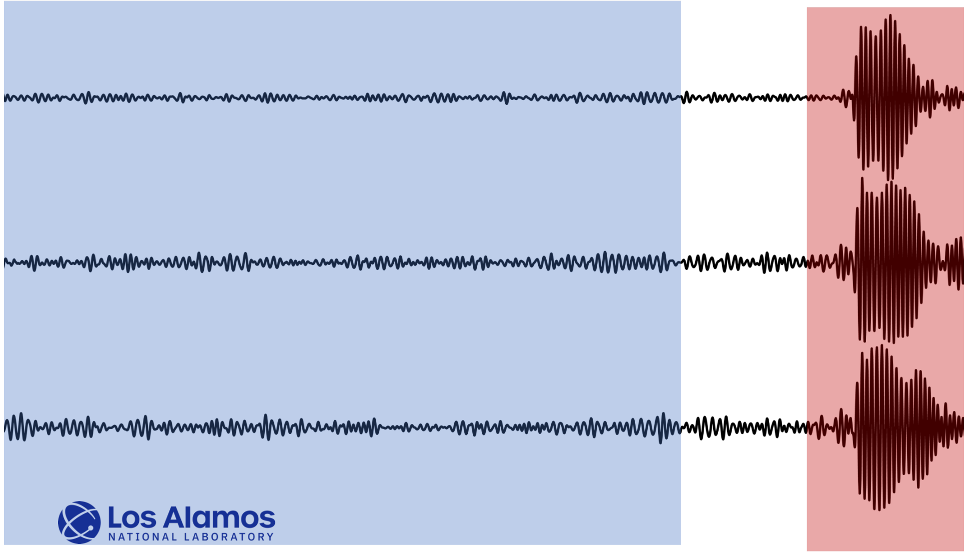
Module 3: Form Competing Data Hypothesis (4/18)



Module 3: Form Competing Data Hypothesis (5/18)

Data is only noise: $\mathcal{H}_0$:

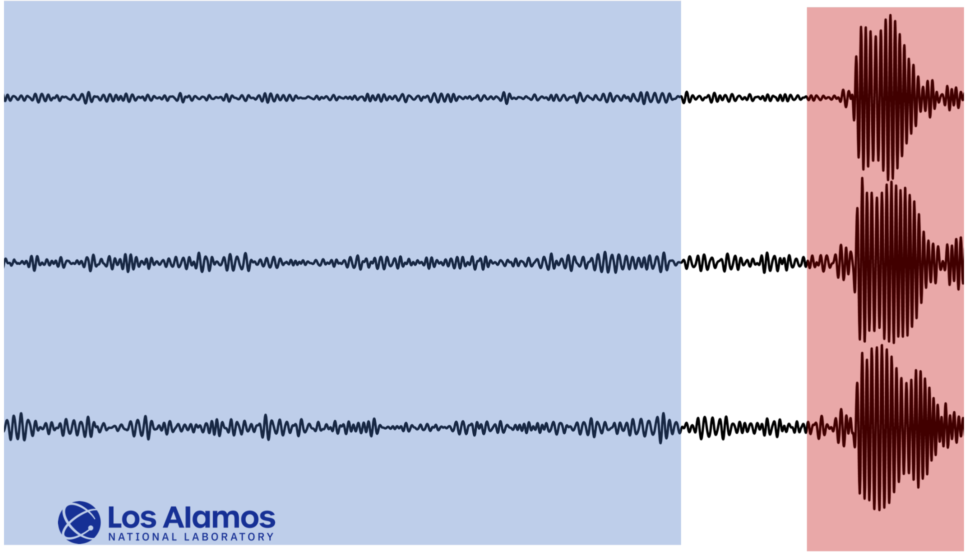
Data includes Rayleigh wave $\mathcal{H}_1$:



Module 3: Form Competing Data Hypothesis (6/18)

Data is only noise: $\mathcal{H}_0 \colon [\boldsymbol{x}_E \quad \boldsymbol{x}_N \quad \boldsymbol{x}_Z] = [\boldsymbol{n}_E \quad \boldsymbol{n}_N \quad \boldsymbol{n}_Z]$

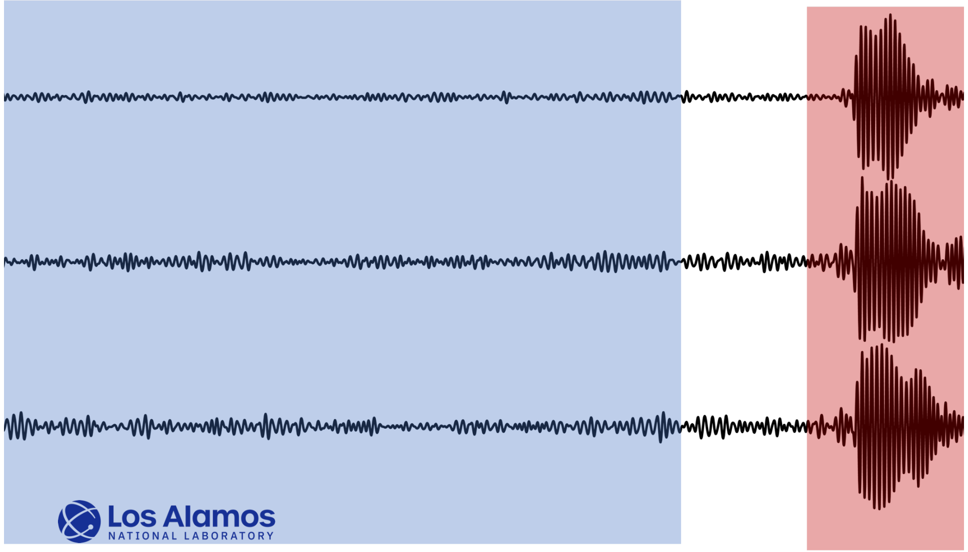
Data includes Rayleigh wave $\mathcal{H}_1 \colon [\boldsymbol{x}_E \quad \boldsymbol{x}_N \quad \boldsymbol{x}_Z] = [\boldsymbol{n}_E \quad \boldsymbol{n}_N \quad \boldsymbol{n}_Z] + [\boldsymbol{s}_E \quad \boldsymbol{s}_N \quad \boldsymbol{s}_Z]$



Module 3: Form Competing Data Hypothesis (7/18)

Data is only noise: $\mathcal{H}_0: [\boldsymbol{x}_E \quad \boldsymbol{x}_N \quad \boldsymbol{x}_Z] = \boxed{[\boldsymbol{n}_E \quad \boldsymbol{n}_N \quad \boldsymbol{n}_Z]}$

Data includes Rayleigh wave $\mathcal{H}_1: [\boldsymbol{x}_E \quad \boldsymbol{x}_N \quad \boldsymbol{x}_Z] = \boxed{[\boldsymbol{n}_E \quad \boldsymbol{n}_N \quad \boldsymbol{n}_Z] + [\boldsymbol{s}_E \quad \boldsymbol{s}_N \quad \boldsymbol{s}_Z]}$

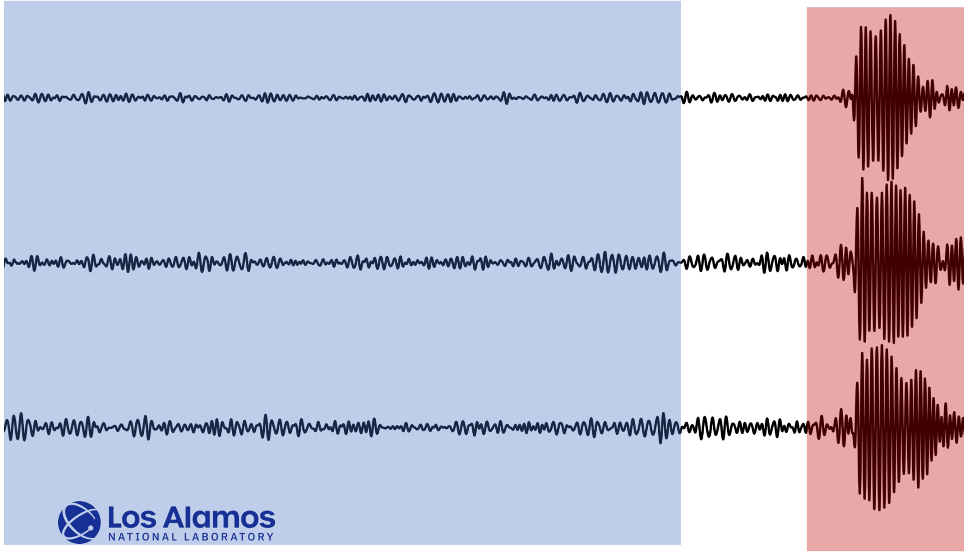


Statistical and deterministic models

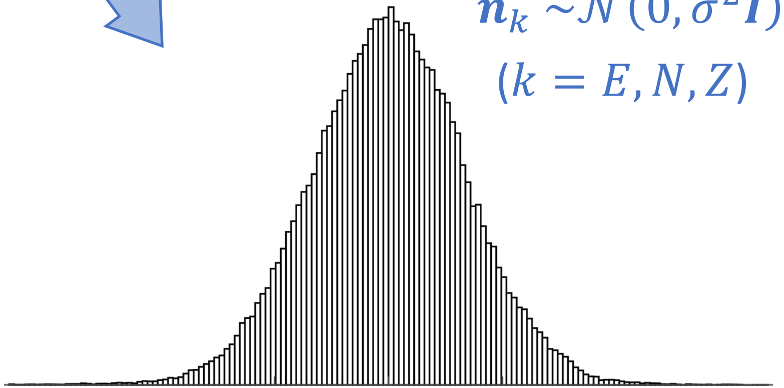
Module 3: Form Competing Data Hypothesis (8/18)

Data is only noise: $\mathcal{H}_0 : [\boldsymbol{x}_E \quad \boldsymbol{x}_N \quad \boldsymbol{x}_Z] = \boxed{[\boldsymbol{n}_E \quad \boldsymbol{n}_N \quad \boldsymbol{n}_Z]}$

Data includes Rayleigh wave $\mathcal{H}_1 : [\boldsymbol{x}_E \quad \boldsymbol{x}_N \quad \boldsymbol{x}_Z] = \boxed{[\boldsymbol{n}_E \quad \boldsymbol{n}_N \quad \boldsymbol{n}_Z] + [\boldsymbol{s}_E \quad \boldsymbol{s}_N \quad \boldsymbol{s}_Z]}$



Statistical and deterministic models



$\boldsymbol{n}_k \sim \mathcal{N}(0, \sigma^2 \boldsymbol{I})$

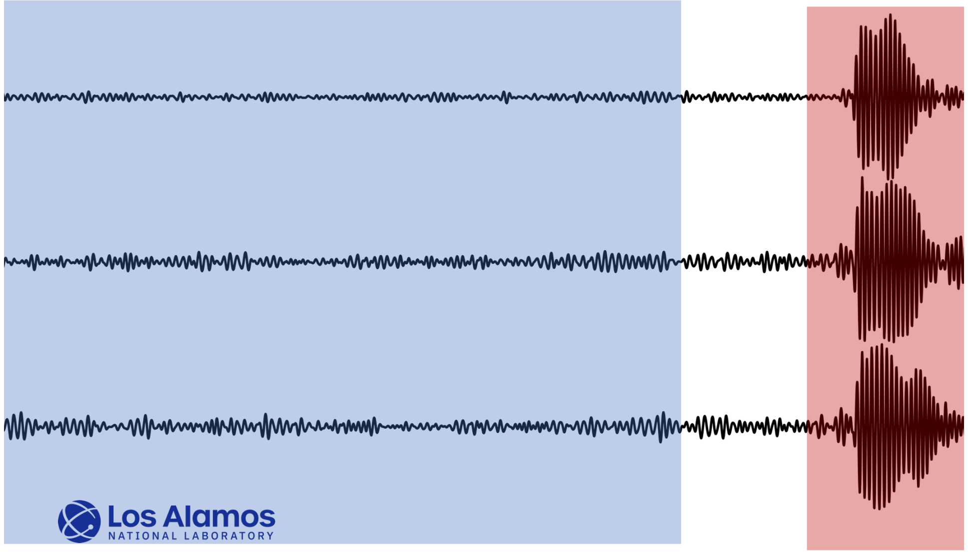
$(k = E, N, Z)$

Los Alamos NATIONAL LABORATORY

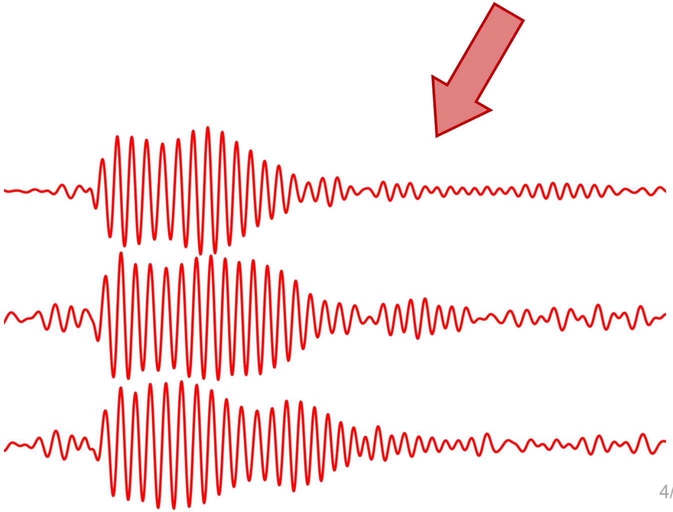
Module 3: Form Competing Data Hypothesis (9/18)

Data is only noise: $\mathcal{H}_0 : [\boldsymbol{x}_E \quad \boldsymbol{x}_N \quad \boldsymbol{x}_Z] = \boxed{[\boldsymbol{n}_E \quad \boldsymbol{n}_N \quad \boldsymbol{n}_Z]}$

Data includes Rayleigh wave $\mathcal{H}_1 : [\boldsymbol{x}_E \quad \boldsymbol{x}_N \quad \boldsymbol{x}_Z] = \boxed{[\boldsymbol{n}_E \quad \boldsymbol{n}_N \quad \boldsymbol{n}_Z] + [\boldsymbol{s}_E \quad \boldsymbol{s}_N \quad \boldsymbol{s}_Z]}$



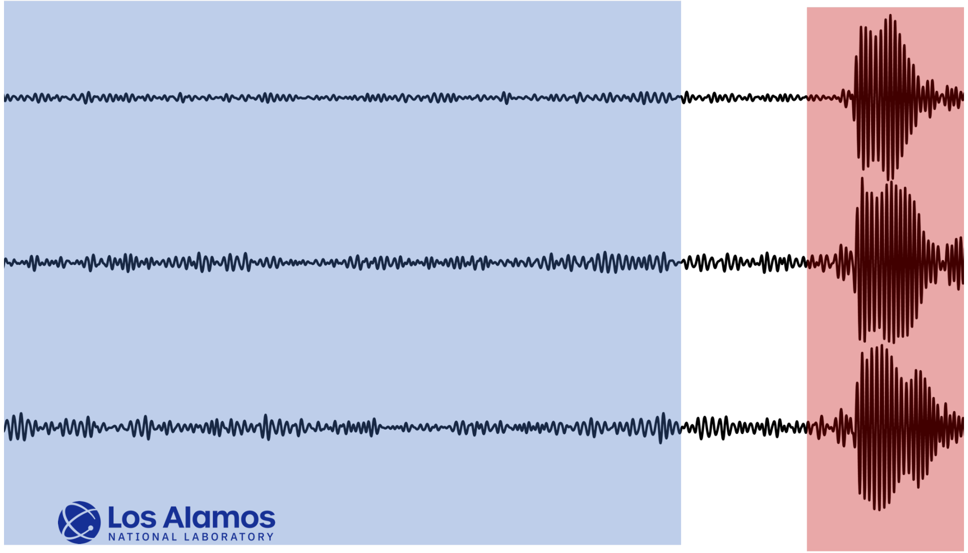
Statistical and deterministic models



Module 3: Form Competing Data Hypothesis (10/18)

Data is only noise: $\mathcal{H}_0: [\boldsymbol{x}_E \quad \boldsymbol{x}_N \quad \boldsymbol{x}_Z] = \boxed{[\boldsymbol{n}_E \quad \boldsymbol{n}_N \quad \boldsymbol{n}_Z]}$

Data includes Rayleigh wave $\mathcal{H}_1: [\boldsymbol{x}_E \quad \boldsymbol{x}_N \quad \boldsymbol{x}_Z] = \boxed{[\boldsymbol{n}_E \quad \boldsymbol{n}_N \quad \boldsymbol{n}_Z] + [\boldsymbol{s}_E \quad \boldsymbol{s}_N \quad \boldsymbol{s}_Z]}$



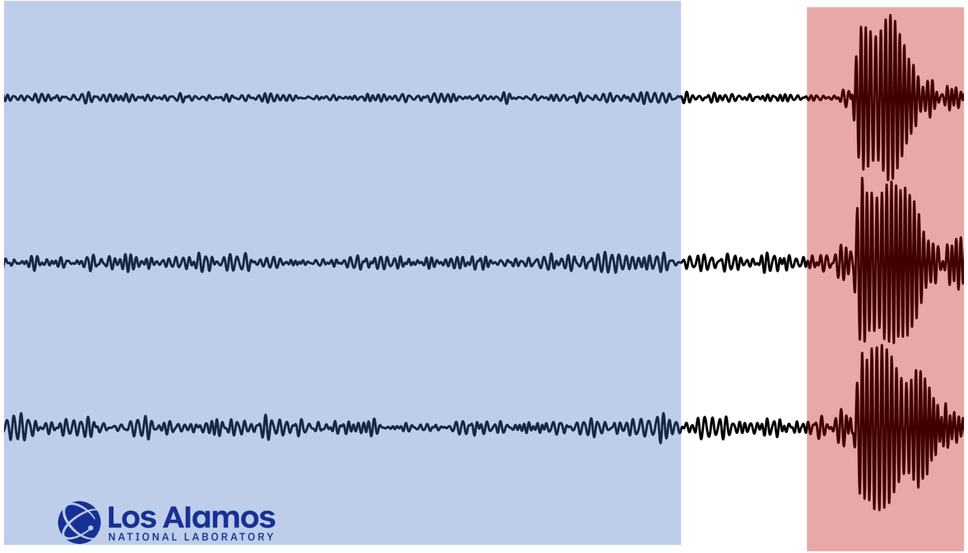
Statistical and deterministic models

$$\boldsymbol{n}_Z \sim \frac{1}{(2\pi\sigma^2)^{\frac{N}{2}}} \exp\left[-\frac{\|\boldsymbol{x}_Z - \boldsymbol{s}_Z\|^2}{2\sigma^2}\right]$$

Module 3: Form Competing Data Hypothesis (11/18)

Data is only noise: $\mathcal{H}_0\colon [\boldsymbol{x}_E \quad \boldsymbol{x}_N \quad \boldsymbol{x}_Z] = \boxed{[\boldsymbol{n}_E \quad \boldsymbol{n}_N \quad \boldsymbol{n}_Z]}$

Data includes Rayleigh wave $\mathcal{H}_1\colon [\boldsymbol{x}_E \quad \boldsymbol{x}_N \quad \boldsymbol{x}_Z] = \boxed{[\boldsymbol{n}_E \quad \boldsymbol{n}_N \quad \boldsymbol{n}_Z] + [\boldsymbol{s}_E \quad \boldsymbol{s}_N \quad \boldsymbol{s}_Z]}$



Statistical and deterministic models

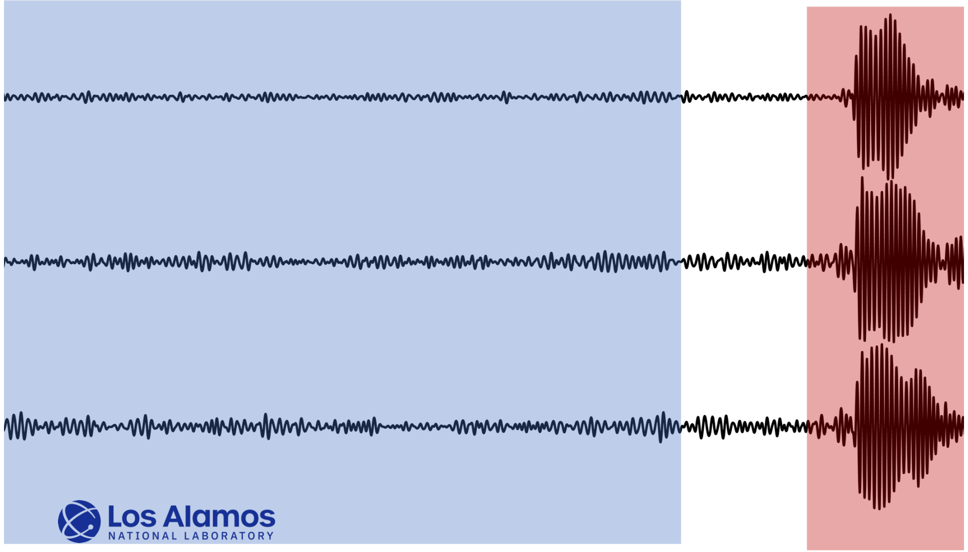
$$\boldsymbol{n}_Z \sim \frac{1}{(2\pi\sigma^2)^{\frac{N}{2}}} \exp\left[-\frac{\|\boldsymbol{x}_Z - \boldsymbol{s}_Z\|^2}{2\sigma^2}\right]$$

The sigma should be a **covariance matrix**, a *scalar* effective degree of freedom parameter accounts for sample covariance

Module 3: Form Competing Data Hypothesis (12/18)

Data is only noise: $\mathcal{H}_0\colon [\boldsymbol{x}_E \quad \boldsymbol{x}_N \quad \boldsymbol{x}_Z] = \boxed{[\boldsymbol{n}_E \quad \boldsymbol{n}_N \quad \boldsymbol{n}_Z]}$

Data includes Rayleigh wave $\mathcal{H}_1\colon [\boldsymbol{x}_E \quad \boldsymbol{x}_N \quad \boldsymbol{x}_Z] = \boxed{[\boldsymbol{n}_E \quad \boldsymbol{n}_N \quad \boldsymbol{n}_Z] + [\boldsymbol{s}_E \quad \boldsymbol{s}_N \quad \boldsymbol{s}_Z]}$



Statistical and deterministic models

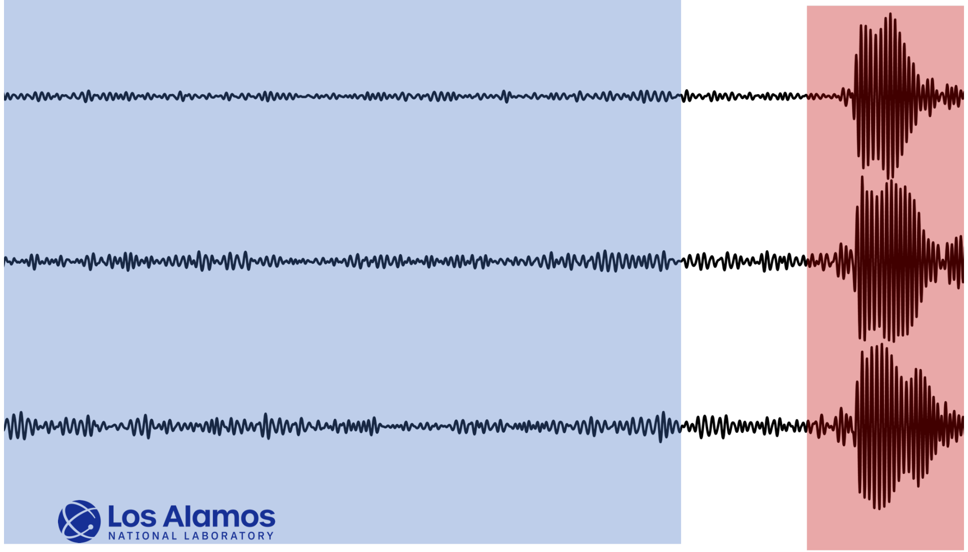
$$\boldsymbol{n}_Z \sim \frac{1}{(2\pi\sigma^2)^{\frac{N}{2}}} \exp\left[-\frac{\|\boldsymbol{x}_Z - \boldsymbol{s}_Z\|^2}{2\sigma^2}\right]$$

Unknowns; includes effective degree of freedom parameter

Module 3: Form Competing Data Hypothesis (13/18)

Data is only noise: $\mathcal{H}_0\colon [\boldsymbol{x}_E \quad \boldsymbol{x}_N \quad \boldsymbol{x}_Z] = \boxed{[\boldsymbol{n}_E \quad \boldsymbol{n}_N \quad \boldsymbol{n}_Z]}$

Data includes Rayleigh wave $\mathcal{H}_1\colon [\boldsymbol{x}_E \quad \boldsymbol{x}_N \quad \boldsymbol{x}_Z] = \boxed{[\boldsymbol{n}_E \quad \boldsymbol{n}_N \quad \boldsymbol{n}_Z] + [\boldsymbol{s}_E \quad \boldsymbol{s}_N \quad \boldsymbol{s}_Z]}$



Statistical and deterministic models

$$\boldsymbol{n}_Z \sim \frac{1}{(2\pi\sigma^2)^{\frac{N}{2}}} \exp\left[-\frac{\|\boldsymbol{x}_Z - \boldsymbol{s}_Z\|^2}{2\sigma^2}\right]$$

Unknowns; includes effective degree of freedom parameter

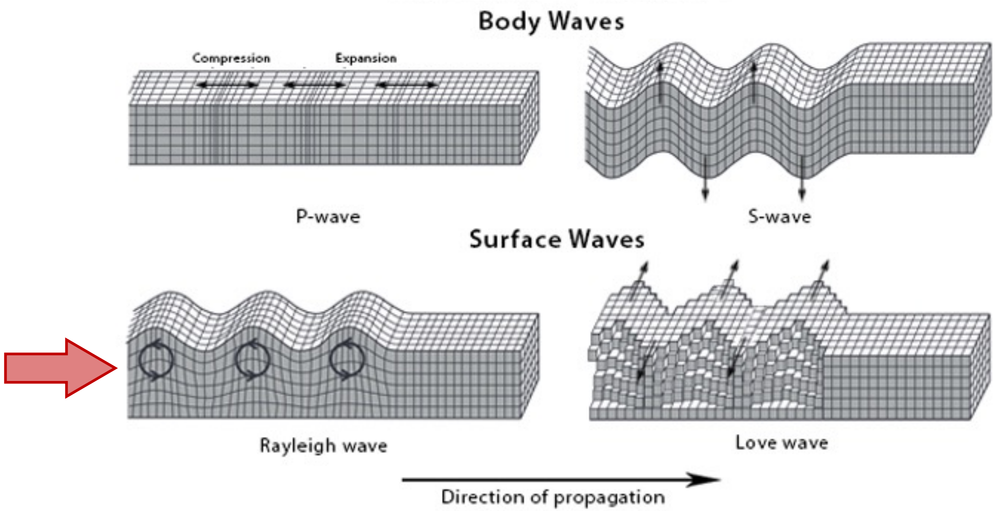
Recall the **Gaussian density** and normalized histograms **match**

Module 3: Form Competing Data Hypothesis (14/18)

Data is only noise: $\mathcal{H}_0: [\boldsymbol{x}_E \quad \boldsymbol{x}_N \quad \boldsymbol{x}_Z] = \boxed{[\boldsymbol{n}_E \quad \boldsymbol{n}_N \quad \boldsymbol{n}_Z]}$

Data includes Rayleigh wave $\mathcal{H}_1: [\boldsymbol{x}_E \quad \boldsymbol{x}_N \quad \boldsymbol{x}_Z] = \boxed{[\boldsymbol{n}_E \quad \boldsymbol{n}_N \quad \boldsymbol{n}_Z] + [\boldsymbol{s}_E \quad \boldsymbol{s}_N \quad \boldsymbol{s}_Z]}$

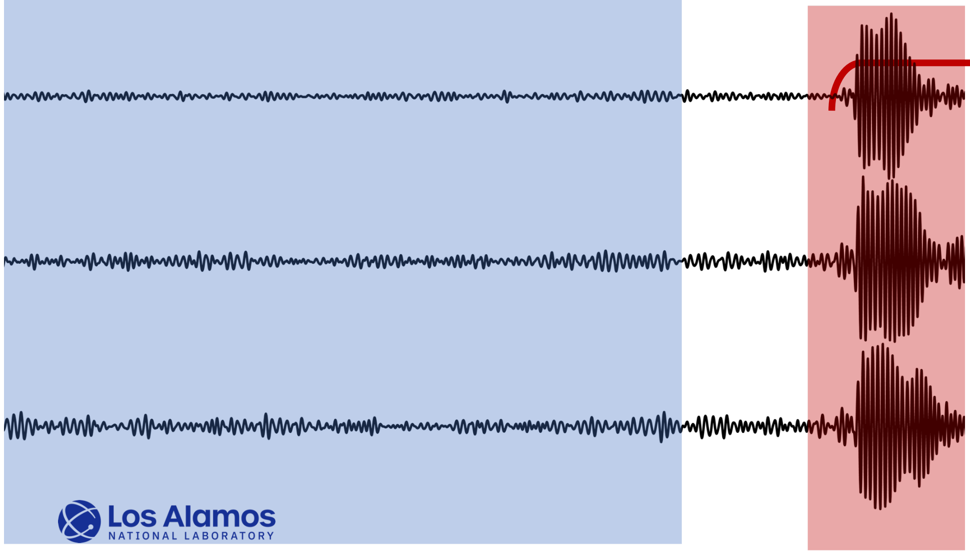
Deterministic Model in Pictures



Module 3: Form Competing Data Hypothesis (15/18)

Data is only noise: $\mathcal{H}_0\colon [\boldsymbol{x}_E \quad \boldsymbol{x}_N \quad \boldsymbol{x}_Z] = \boxed{[\boldsymbol{n}_E \quad \boldsymbol{n}_N \quad \boldsymbol{n}_Z]}$

Data includes Rayleigh wave $\mathcal{H}_1\colon [\boldsymbol{x}_E \quad \boldsymbol{x}_N \quad \boldsymbol{x}_Z] = \boxed{[\boldsymbol{n}_E \quad \boldsymbol{n}_N \quad \boldsymbol{n}_Z] + [\boldsymbol{s}_E \quad \boldsymbol{s}_N \quad \boldsymbol{s}_Z]}$



Statistical and deterministic models

$$x_Z(\mathbb{r},\omega) \propto \sum_n \frac{r_2(\mathfrak{z})}{8cUI_1}\sqrt{\frac{2}{\pi k_n \mathbb{r}}}\exp\left[j\left(k_n\mathbb{r}+\frac{\pi}{4}\right)\right]\{\blacksquare\}$$

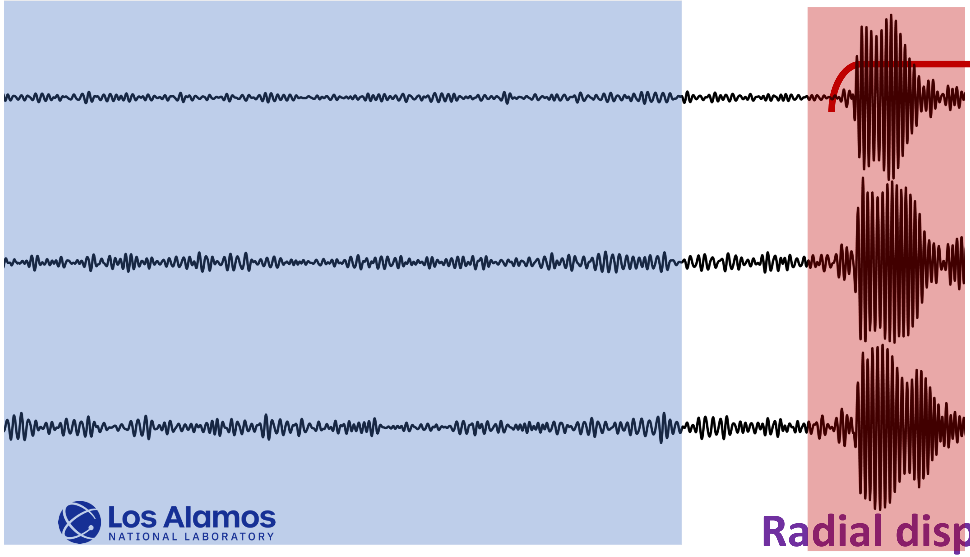
$$x_R(\mathbb{r},\omega) \propto \sum_n \frac{r_1(\mathfrak{z})}{8cUI_1}\sqrt{\frac{2}{\pi k_n \mathbb{r}}}\exp\left[j\left(k_n\mathbb{r}-\frac{\pi}{4}\right)\right]\{\blacksquare\}$$

Aki and Richards (Eq. 7.150-7.151)

Module 3: Form Competing Data Hypothesis (16/18)

Data is only noise: $\mathcal{H}_0\colon [\boldsymbol{x}_E \quad \boldsymbol{x}_N \quad \boldsymbol{x}_Z] = \boxed{[\boldsymbol{n}_E \quad \boldsymbol{n}_N \quad \boldsymbol{n}_Z]}$

Data includes Rayleigh wave $\mathcal{H}_1\colon [\boldsymbol{x}_E \quad \boldsymbol{x}_N \quad \boldsymbol{x}_Z] = \boxed{[\boldsymbol{n}_E \quad \boldsymbol{n}_N \quad \boldsymbol{n}_Z] + [\boldsymbol{s}_E \quad \boldsymbol{s}_N \quad \boldsymbol{s}_Z]}$



Statistical and deterministic models

$$x_Z(\mathbb{r},\omega) \propto \sum_n \frac{r_2(\mathfrak{z})}{8cUI_1}\sqrt{\frac{2}{\pi k_n \mathbb{r}}}\exp\left[j\left(k_n\mathbb{r}+\frac{\pi}{4}\right)\right]\{\blacksquare\}$$

$$x_R(\mathbb{r},\omega) \propto \sum_n \frac{r_1(\mathfrak{z})}{8cUI_1}\sqrt{\frac{2}{\pi k_n \mathbb{r}}}\exp\left[j\left(k_n\mathbb{r}-\frac{\pi}{4}\right)\right]\{\blacksquare\}$$

Radial displacement Amplitudes 90 phase advance

Module 3: Form Competing Data Hypothesis (17/18)

Data is only noise: $\mathcal{H}_0: [\boldsymbol{x}_E \quad \boldsymbol{x}_N \quad \boldsymbol{x}_Z] = \boxed{[\boldsymbol{n}_E \quad \boldsymbol{n}_N \quad \boldsymbol{n}_Z]}$

Data includes Rayleigh wave $\mathcal{H}_1: [\boldsymbol{x}_E \quad \boldsymbol{x}_N \quad \boldsymbol{x}_Z] = \boxed{[\boldsymbol{n}_E \quad \boldsymbol{n}_N \quad \boldsymbol{n}_Z] + [\boldsymbol{s}_E \quad \boldsymbol{s}_N \quad \boldsymbol{s}_Z]}$

In Words

The previous slides model the noise as Gaussian, but with **unknown** noise variance that we can estimate.

If data contain a Rayleigh wave, the previous slide **also** models the radial observed displacement (or velocity) as quasi-proportional to the vertical component of the observed displacement, after a 90-degree phase delay. The data contain **unknown** source location and amplitude proportionality constant.

Module 3: Form Competing Data Hypothesis (18/18)

Data is only noise: $\mathcal{H}_0\colon [\boldsymbol{x}_E \quad \boldsymbol{x}_N \quad \boldsymbol{x}_Z] = \boxed{[\boldsymbol{n}_E \quad \boldsymbol{n}_N \quad \boldsymbol{n}_Z]}$

Data includes Rayleigh wave $\mathcal{H}_1\colon [\boldsymbol{x}_E \quad \boldsymbol{x}_N \quad \boldsymbol{x}_Z] = \boxed{[\boldsymbol{n}_E \quad \boldsymbol{n}_N \quad \boldsymbol{n}_Z] + [\boldsymbol{s}_E \quad \boldsymbol{s}_N \quad \boldsymbol{s}_Z]}$

Combine radial rotation and phase advance information:

$$\boldsymbol{x}_R = A\cos\alpha\,\boldsymbol{x}_N + B\sin\alpha\;\boldsymbol{x}_E \equiv \theta_1\boldsymbol{x}_N + \theta_2\,\boldsymbol{x}_E$$

90^{o} Phase-Advance $\{\boldsymbol{x}_Z\} \equiv \mathcal{J}[\boldsymbol{x}_Z] \propto \boldsymbol{x}_R$

Hilbert transform

Module 4:
Build Test Statistic

Module 4: Build Test Statistic (1/10)

Data is only noise $\mathcal{H}_0\!: \ \mathcal{J}[\boldsymbol{x}_Z] = \boldsymbol{n} \sim \mathcal{N}(\boldsymbol{0}, \sigma^2 \boldsymbol{I})$

Data includes Rayleigh wave

$$\mathcal{H}_1\!: \ \mathcal{J}[\boldsymbol{x}_Z] = [\boldsymbol{x}_E \quad \boldsymbol{x}_N]\boldsymbol{\theta} + \boldsymbol{n} \sim \mathcal{N}([\boldsymbol{x}_E \quad \boldsymbol{x}_N]\boldsymbol{\theta}, \sigma^2 \boldsymbol{I})$$

$$\equiv \boldsymbol{U\theta} + \boldsymbol{n} \sim \mathcal{N}(\boldsymbol{U\theta}, \sigma^2 \boldsymbol{I})$$

Module 4: Build Test Statistic (2/10)

Data is only noise $\mathcal{H}_0$: $\mathcal{J}[\boldsymbol{x}_Z] = \boldsymbol{n} \sim \mathcal{N}(\boldsymbol{0}, \sigma^2 \boldsymbol{I})$

Data includes Rayleigh wave $\mathcal{H}_1$: $\mathcal{J}[\boldsymbol{x}_Z] = [\boldsymbol{x}_E \quad \boldsymbol{x}_N]\boldsymbol{\theta} + \boldsymbol{n} \sim \mathcal{N}([\boldsymbol{x}_E \quad \boldsymbol{x}_N]\boldsymbol{\theta}, \sigma^2 \boldsymbol{I})$

$$\equiv \boldsymbol{U\theta} + \boldsymbol{n} \sim \mathcal{N}(\boldsymbol{U\theta}, \sigma^2 \boldsymbol{I})$$

Noise is also transformed, but rotated and Hilbert transformed Gaussian noise is still Gaussian noise, so we can write it as $\boldsymbol{n}$.

Module 4: Build Test Statistic (3/10)

Data is only noise $\mathcal{H}_0\text{:}\ \mathcal{J}[\boldsymbol{x}_Z] = \boldsymbol{n} \sim \mathcal{N}(\boldsymbol{0}, \sigma^2\boldsymbol{I})$

Data includes Rayleigh wave $\mathcal{H}_1\text{:}\ \mathcal{J}[\boldsymbol{x}_Z] = [\boldsymbol{x}_E \quad \boldsymbol{x}_N]\boldsymbol{\theta} + \boldsymbol{n} \sim \mathcal{N}([\boldsymbol{x}_E \quad \boldsymbol{x}_N]\boldsymbol{\theta}, \sigma^2\boldsymbol{I})$

$$\equiv \boldsymbol{U\theta} + \boldsymbol{n} \sim \mathcal{N}(\boldsymbol{U\theta}, \sigma^2\boldsymbol{I})$$

$$\mathrm{GLR} = \max_{\sigma^2\ \boldsymbol{\theta}} \left\{ \frac{1}{(2\pi\sigma^2)^{\frac{N}{2}}} \exp\left[-\frac{\|\mathcal{J}[\boldsymbol{x}_Z] - \boldsymbol{U\theta}\|^2}{2\sigma^2}\right] \right\} \Bigg/ \max_{\sigma^2} \left\{ \frac{1}{(2\pi\sigma^2)^{\frac{N}{2}}} \exp\left[-\frac{\|\mathcal{J}[\boldsymbol{x}_Z]\|^2}{2\sigma^2}\right] \right\}$$

Module 4: Build Test Statistic (4/10)

Data is only noise $\mathcal{H}_0\!: \ \mathcal{J}[\boldsymbol{x}_Z] = \boldsymbol{n} \sim \mathcal{N}(\boldsymbol{0}, \sigma^2\boldsymbol{I})$

Data includes
Rayleigh wave

$$\mathcal{H}_1\!: \ \mathcal{J}[\boldsymbol{x}_Z] = [\boldsymbol{x}_E \quad \boldsymbol{x}_N]\boldsymbol{\theta} + \boldsymbol{n} \sim \mathcal{N}([\boldsymbol{x}_E \quad \boldsymbol{x}_N]\boldsymbol{\theta}, \sigma^2\boldsymbol{I})$$

$$\equiv \boldsymbol{U\theta} + \boldsymbol{n} \sim \mathcal{N}(\boldsymbol{U\theta}, \sigma^2\boldsymbol{I})$$

$$\mathrm{GLR} = \max_{\sigma^2\,\boldsymbol{\theta}} \left\{ \frac{1}{(2\pi\sigma^2)^{\frac{N}{2}}} \exp\left[-\frac{\|\mathcal{J}[\boldsymbol{x}_Z] - \boldsymbol{U\theta}\|^2}{2\sigma^2} \right] \right\} \Bigg/ \max_{\sigma^2} \left\{ \frac{1}{(2\pi\sigma^2)^{\frac{N}{2}}} \exp\left[-\frac{\|\mathcal{J}[\boldsymbol{x}_Z]\|^2}{2\sigma^2} \right] \right\}$$

Maximize over unknown parameters in the numerator and denominator

Module 4: Build Test Statistic (5/10)

Data is only noise $\mathcal{H}_0\!: \ \mathcal{J}[\boldsymbol{x}_Z] = \boldsymbol{n} \sim \mathcal{N}(\boldsymbol{0}, \sigma^2 \boldsymbol{I})$

Data includes Rayleigh wave $\mathcal{H}_1\!: \ \mathcal{J}[\boldsymbol{x}_Z] = [\boldsymbol{x}_E \quad \boldsymbol{x}_N]\boldsymbol{\theta} + \boldsymbol{n} \sim \mathcal{N}([\boldsymbol{x}_E \quad \boldsymbol{x}_N]\boldsymbol{\theta}, \sigma^2 \boldsymbol{I})$

$$\equiv \boldsymbol{U\theta} + \boldsymbol{n} \sim \mathcal{N}(\boldsymbol{U\theta}, \sigma^2 \boldsymbol{I})$$

$$\mathrm{GLR} = \max_{\sigma^2\ \boldsymbol{\theta}} \left\{ \frac{1}{(2\pi\sigma^2)^{\frac{N}{2}}} \exp\left[-\frac{\|\mathcal{J}[\boldsymbol{x}_Z] - \boldsymbol{U\theta}\|^2}{2\sigma^2} \right] \right\} \Big/ \max_{\sigma^2} \left\{ \frac{1}{(2\pi\sigma^2)^{\frac{N}{2}}} \exp\left[-\frac{\|\mathcal{J}[\boldsymbol{x}_Z]\|^2}{2\sigma^2} \right] \right\}$$

Maximize over unknown parameters in the numerator and denominator

Call $\mathcal{J}[\boldsymbol{x}_Z]$ the (post-processed) variable $\boldsymbol{x}_Z$ hereon to make notation clearer

Los Alamos NATIONAL LABORATORY

Module 4: Build Test Statistic (6/10)

Data is only noise $\mathcal{H}_0\colon\ \mathcal{J}[\boldsymbol{x}_Z] = \boldsymbol{n} \sim \mathcal{N}(\boldsymbol{0}, \sigma^2 \boldsymbol{I})$

Data includes Rayleigh wave $\mathcal{H}_1\colon\ \mathcal{J}[\boldsymbol{x}_Z] = [\boldsymbol{x}_E \quad \boldsymbol{x}_N]\boldsymbol{\theta} + \boldsymbol{n} \sim \mathcal{N}([\boldsymbol{x}_E \quad \boldsymbol{x}_N]\boldsymbol{\theta}, \sigma^2 \boldsymbol{I})$

$$\equiv \boldsymbol{U\theta} + \boldsymbol{n} \sim \mathcal{N}(\boldsymbol{U\theta}, \sigma^2 \boldsymbol{I})$$

Shorthand

$$\left(\frac{N-2}{2}\right)\left(\mathrm{GLR}^{2/N} - 1\right) = \gamma \quad \Rightarrow \quad \overbrace{\left(\frac{N-2}{2}\right) \frac{\boldsymbol{x}_Z^{\mathrm{T}} \boldsymbol{U}\left(\boldsymbol{U}^{\mathrm{T}}\boldsymbol{U}\right)^{-1}\boldsymbol{U}^{\mathrm{T}}\boldsymbol{x}_Z}{\boldsymbol{x}_Z^{\mathrm{T}}(\boldsymbol{I} - \boldsymbol{U}(\boldsymbol{U}^{\mathrm{T}}\boldsymbol{U})^{-1}\boldsymbol{U}^{\mathrm{T}})\boldsymbol{x}_Z}}^{s(\boldsymbol{x}_Z)} = \gamma$$

samples in processing window

Module 4: Build Test Statistic (7/10)

Data is only noise $\mathcal{H}_0$: $\mathcal{J}[\boldsymbol{x}_Z] = \boldsymbol{n} \sim \mathcal{N}(\boldsymbol{0}, \sigma^2 \boldsymbol{I})$

Data includes Rayleigh wave $\mathcal{H}_1$: $\mathcal{J}[\boldsymbol{x}_Z] = [\boldsymbol{x}_E \quad \boldsymbol{x}_N]\boldsymbol{\theta} + \boldsymbol{n} \sim \mathcal{N}([\boldsymbol{x}_E \quad \boldsymbol{x}_N]\boldsymbol{\theta}, \sigma^2 \boldsymbol{I})$

$$\equiv \boldsymbol{U}\boldsymbol{\theta} + \boldsymbol{n} \sim \mathcal{N}(\boldsymbol{U}\boldsymbol{\theta}, \sigma^2 \boldsymbol{I})$$

Shorthand

$$\left(\frac{N-2}{2}\right)\left(\mathrm{GLR}^{2/N} - 1\right) = \gamma \quad \Rightarrow \quad \overbrace{\left(\frac{N-2}{2}\right) \frac{\boldsymbol{x}_Z^{\mathrm{T}} \boldsymbol{U}\left(\boldsymbol{U}^{\mathrm{T}}\boldsymbol{U}\right)^{-1} \boldsymbol{U}^{\mathrm{T}} \boldsymbol{x}_Z}{\boldsymbol{x}_Z^{\mathrm{T}}\left(\boldsymbol{I} - \boldsymbol{U}(\boldsymbol{U}^{\mathrm{T}}\boldsymbol{U})^{-1}\boldsymbol{U}^{\mathrm{T}}\right)\boldsymbol{x}_Z}}^{s(\boldsymbol{x}_Z)} = \gamma$$

samples in processing window

$$\sim \mathcal{F}_{D1,D2}(0) \text{ under } \mathcal{H}_0$$

$$\sim \mathcal{F}_{D1,D2}\left(\frac{\boldsymbol{\theta}^{\mathrm{T}}\boldsymbol{\theta}}{\sigma^2}\right) \text{ under } \mathcal{H}_1$$

Module 4: Build Test Statistic (8/10)

The distribution of the detection statistic $s(\boldsymbol{x}_Z)$ informs the detector algorithm what density the histogram will match, **and** it will allow you to set a threshold for declaring that you <u>have</u> or <u>have not</u> detected a Rayleigh wave.

Shorthand

$$\left(\frac{N-2}{2}\right)\left(\mathrm{GLR}^{2/N}-1\right)=\gamma \quad \Rightarrow \quad \overbrace{\left(\frac{N-2}{2}\right)\frac{\boldsymbol{x}_Z^{\mathrm{T}}\boldsymbol{U}\left(\boldsymbol{U}^{\mathrm{T}}\boldsymbol{U}\right)^{-1}\boldsymbol{U}^{\mathrm{T}}\boldsymbol{x}_Z}{\boldsymbol{x}_Z^{\mathrm{T}}\left(\boldsymbol{I}-\boldsymbol{U}\left(\boldsymbol{U}^{\mathrm{T}}\boldsymbol{U}\right)^{-1}\boldsymbol{U}^{\mathrm{T}}\right)\boldsymbol{x}_Z}}^{s(\boldsymbol{x}_Z)}=\gamma$$

samples in processing window

$\sim \mathcal{F}_{D1,D2}(0)$ under $\mathcal{H}_0$

$\sim \mathcal{F}_{D1,D2}\left(\frac{\boldsymbol{\theta}^{\mathrm{T}}\boldsymbol{\theta}}{\sigma^2}\right)$ under $\mathcal{H}_1$

Module 4: Build Test Statistic (9/10)

The distribution of the detection statistic $s(\boldsymbol{x}_Z)$ informs the detector algorithm what density the histogram will match, **and** it will allow you to set a threshold for declaring that you have or have not detected a Rayleigh wave.

Important: this detection statistic is *not Gaussian*. The data $\boldsymbol{x}_Z$ input to $s(\boldsymbol{x}_Z)$ *is* Gaussian. The detection statistic *output* is *not*.

$$\overbrace{\left(\frac{N-2}{2}\right)\frac{\boldsymbol{x}_Z^{\mathrm{T}}\boldsymbol{U}\left(\boldsymbol{U}^{\mathrm{T}}\boldsymbol{U}\right)^{-1}\boldsymbol{U}^{\mathrm{T}}\boldsymbol{x}_Z}{\boldsymbol{x}_Z^{\mathrm{T}}(\boldsymbol{I}-\boldsymbol{U}(\boldsymbol{U}^{\mathrm{T}}\boldsymbol{U})^{-1}\boldsymbol{U}^{\mathrm{T}})\boldsymbol{x}_Z}}^{s(\boldsymbol{x}_Z)} = \gamma$$

$\sim \mathcal{F}_{D1,D2}(0)$ under $\mathcal{H}_0$ **Not Gaussian**

$\sim \mathcal{F}_{D1,D2}\left(\frac{\boldsymbol{\theta}^{\mathrm{T}}\boldsymbol{\theta}}{\sigma^2}\right)$ under $\mathcal{H}_1$ **Not Gaussian**

Module 4: Build Test Statistic (10/10)

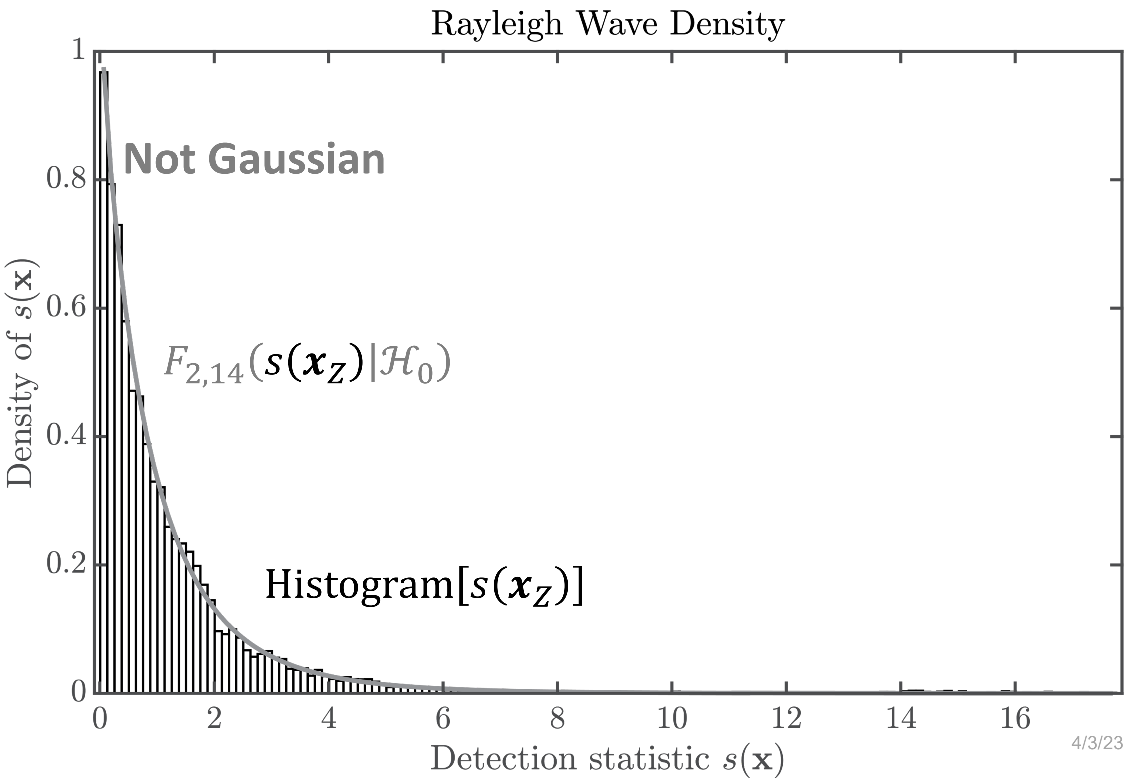
Test $\mathcal{F}$**-distribution** against ∼1 hr of the target data

$$\underbrace{\left(\frac{N-2}{2}\right)\frac{\boldsymbol{x}_Z^{\mathrm{T}}\boldsymbol{U}\left(\boldsymbol{U}^{\mathrm{T}}\boldsymbol{U}\right)^{-1}\boldsymbol{U}^{\mathrm{T}}\boldsymbol{x}_Z}{\boldsymbol{x}_Z^{\mathrm{T}}(\boldsymbol{I}-\boldsymbol{U}(\boldsymbol{U}^{\mathrm{T}}\boldsymbol{U})^{-1}\boldsymbol{U}^{\mathrm{T}})\boldsymbol{x}_Z}}_{s(\boldsymbol{x}_Z)}$$

Density Estimation

- Compute $s(\boldsymbol{x}_Z)$ at every time sample in three channel data
- Exclude extreme quantiles from detection statistic
- Compare normalized histogram against theoretical density

Select distributional parameters to minimize density-histogram mismatch



Module 5:
Compute Test Statistic

Status Update: What we've Done Already

Single modality detection: terminology, concepts

$\mathcal{H}_0$: $\boldsymbol{x} =$ noise + background

$\mathcal{H}_1$: $\boldsymbol{x} =$ noise + background + signal

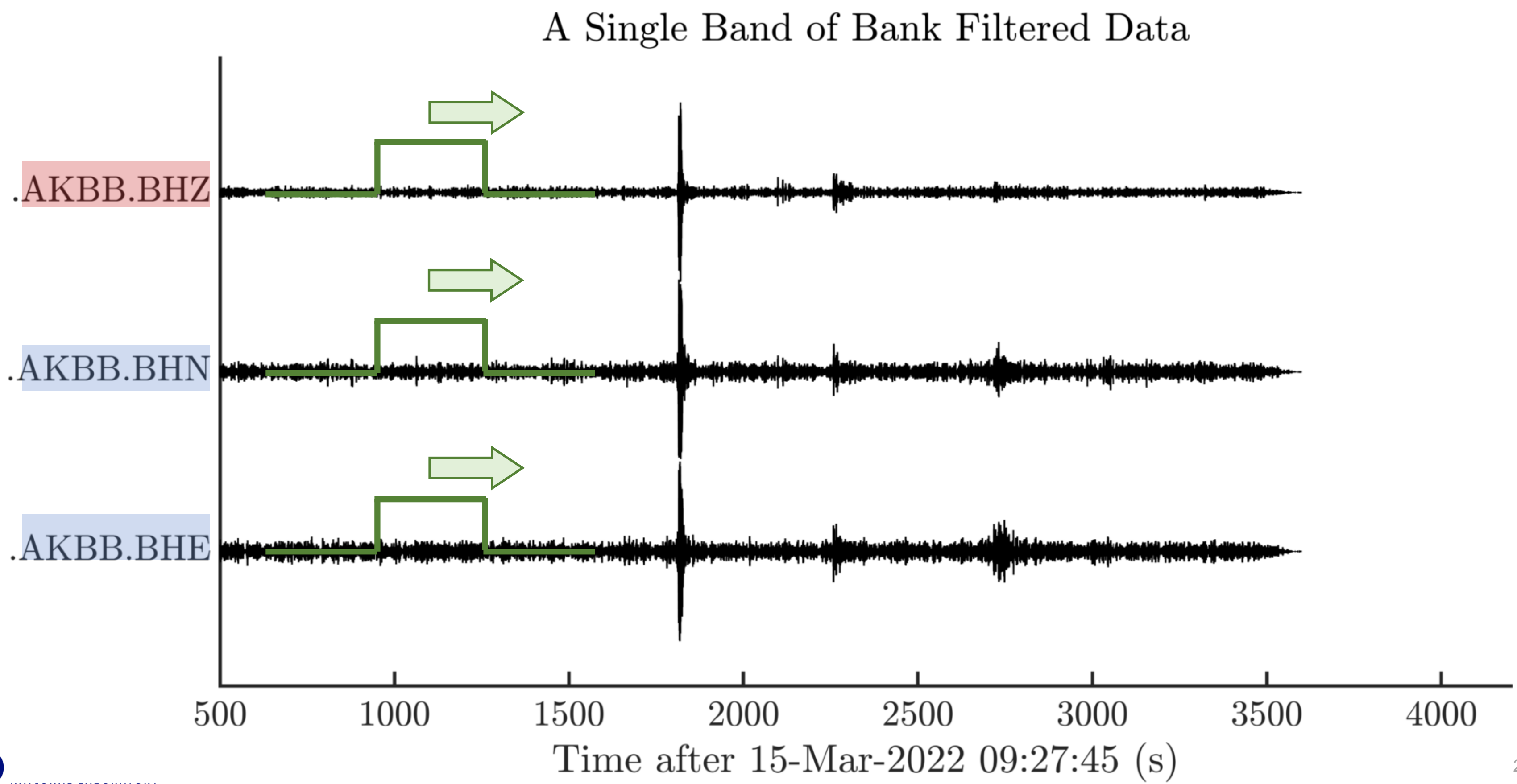
$$\frac{\widehat{\mathrm{PDF}}(\boldsymbol{x}|\mathcal{H}_1)}{\widehat{\mathrm{PDF}}(\boldsymbol{x}|\mathcal{H}_0)} = s(\boldsymbol{x}) \underset{\mathcal{H}_0}{\overset{\mathcal{H}_1}{\gtrless}} t \qquad \begin{array}{l} \mathcal{H}_1 \rightarrow u = 1 \\ \mathcal{H}_0 \rightarrow u = 0 \end{array}$$

Source | **Data Models** | **GLR** | **Detector** | **Decision**

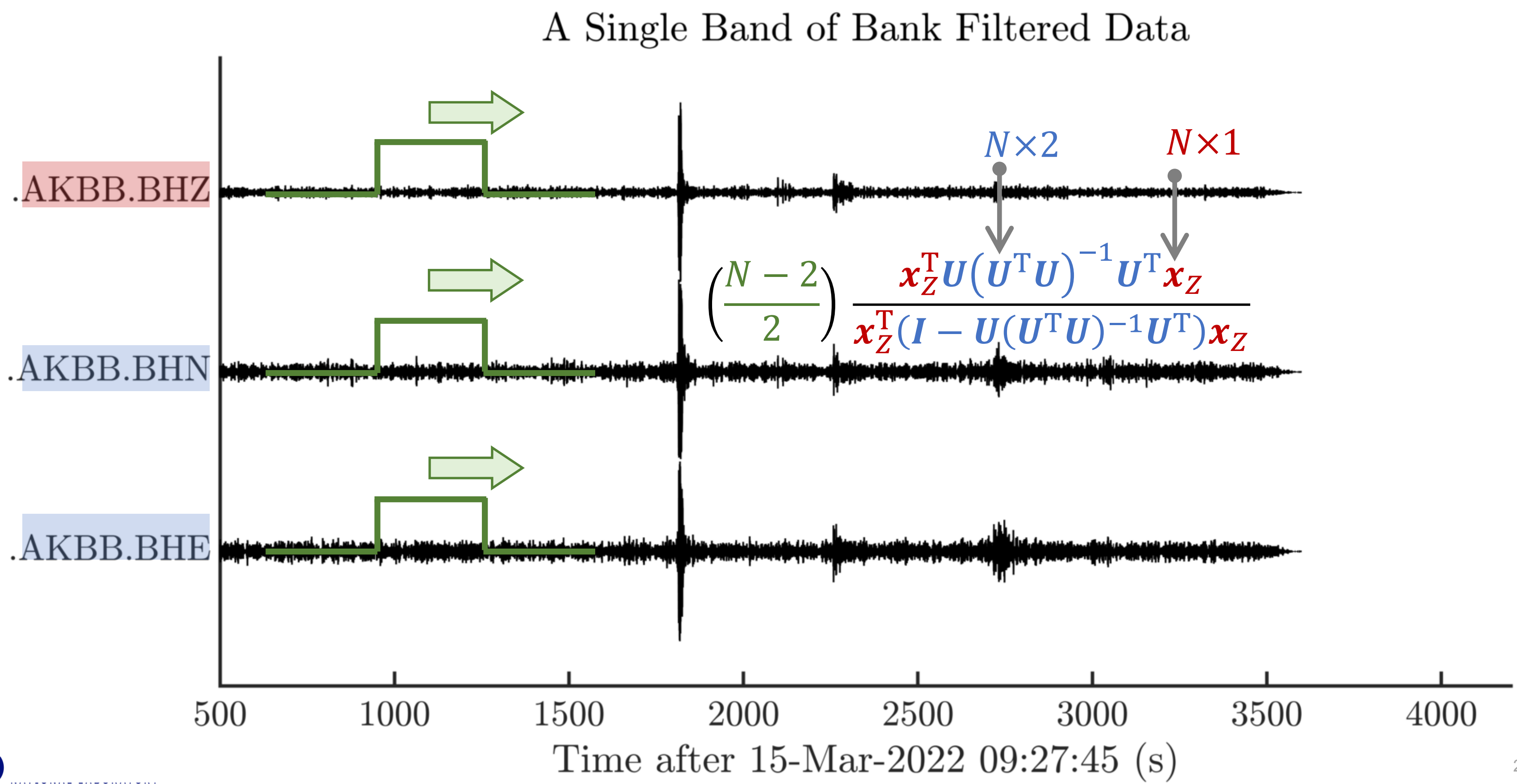
Density processing: estimate distributional parameters that shape density and histograms

Data pre-processing: detrend, taper and bandpass filter (possibly in a filter bank)

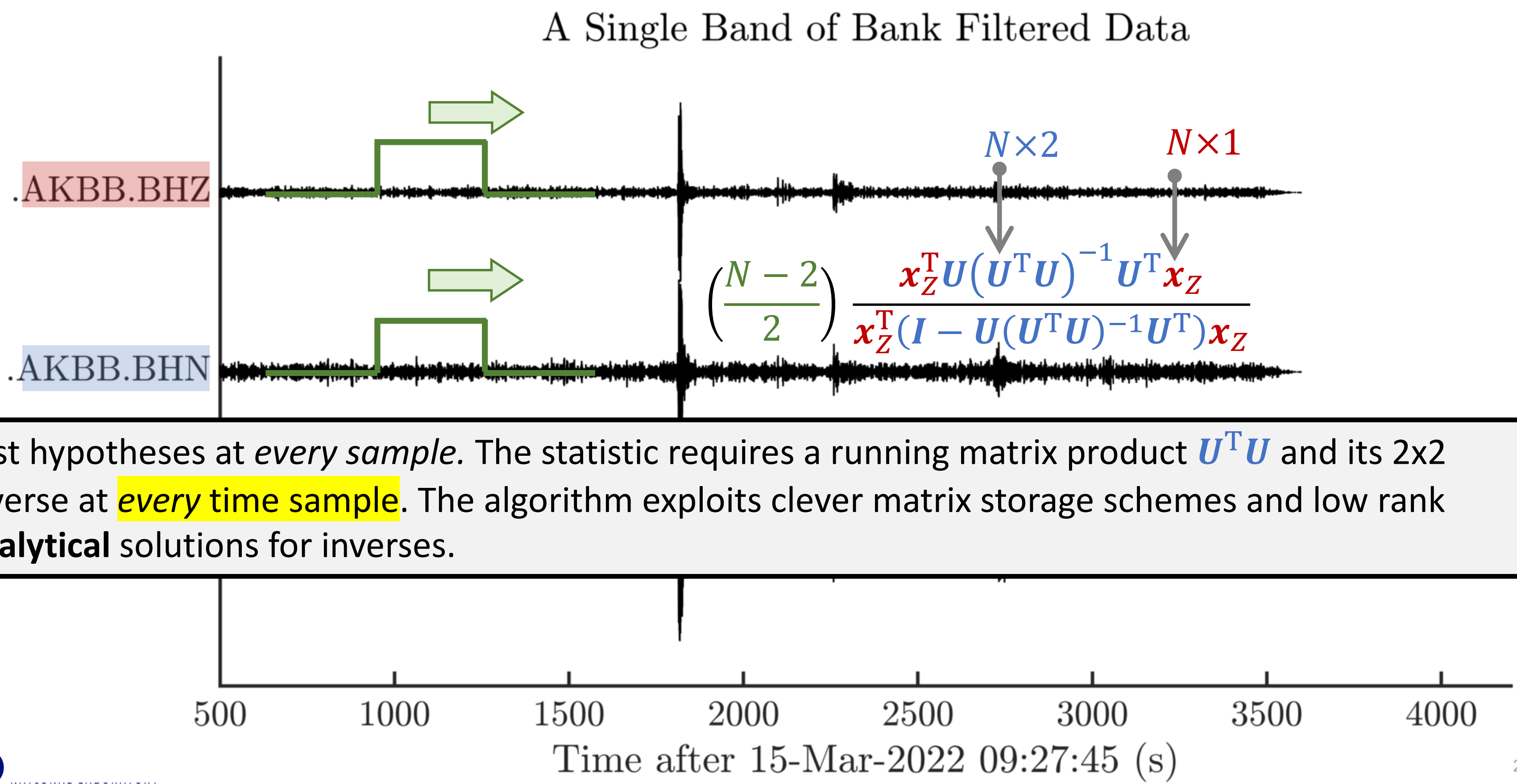
Module 5: Compute Test Statistic (1/10)



Module 5: Compute Test Statistic (2/10)



Module 5: Compute Test Statistic (3/10)



Test hypotheses at *every sample.* The statistic requires a running matrix product $\boldsymbol{U}^{\mathrm{T}}\boldsymbol{U}$ and its 2x2 inverse at *every* time sample. The algorithm exploits clever matrix storage schemes and low rank **analytical** solutions for inverses.

Module 5: Compute Test Statistic (4/10)

$$\left(\frac{N-2}{2}\right)\frac{\boldsymbol{x}_Z^{\mathrm{T}}\boldsymbol{U}\left(\boldsymbol{U}^{\mathrm{T}}\boldsymbol{U}\right)^{-1}\boldsymbol{U}^{\mathrm{T}}\boldsymbol{x}_Z}{\boldsymbol{x}_Z^{\mathrm{T}}(\boldsymbol{I}-\boldsymbol{U}(\boldsymbol{U}^{\mathrm{T}}\boldsymbol{U})^{-1}\boldsymbol{U}^{\mathrm{T}})\boldsymbol{x}_Z}$$

$$\boldsymbol{U}^{\mathrm{T}}\boldsymbol{U}=\begin{bmatrix}\boldsymbol{U}_1^T\boldsymbol{U}_1 & \boldsymbol{U}_1^T\boldsymbol{U}_2\\ \boldsymbol{U}_1^T\boldsymbol{U}_2 & \boldsymbol{U}_2^T\boldsymbol{U}_2\end{bmatrix}$$

This product is a 2x2 matrix

Module 5: Compute Test Statistic (5/10)

$$\left(\frac{N-2}{2}\right)\frac{\boldsymbol{x}_Z^{\mathrm{T}}\boldsymbol{U}\left(\boldsymbol{U}^{\mathrm{T}}\boldsymbol{U}\right)^{-1}\boldsymbol{U}^{\mathrm{T}}\boldsymbol{x}_Z}{\boldsymbol{x}_Z^{\mathrm{T}}\left(\boldsymbol{I}-\boldsymbol{U}(\boldsymbol{U}^{\mathrm{T}}\boldsymbol{U})^{-1}\boldsymbol{U}^{\mathrm{T}}\right)\boldsymbol{x}_Z}$$

$$\boldsymbol{A}=\boldsymbol{U}^{\mathrm{T}}\boldsymbol{U}=\begin{bmatrix}\boldsymbol{U}_1^T\boldsymbol{U}_1 & \boldsymbol{U}_1^T\boldsymbol{U}_2\\ \boldsymbol{U}_1^T\boldsymbol{U}_2 & \boldsymbol{U}_2^T\boldsymbol{U}_2\end{bmatrix}$$

```
%first compute U'*U: Matrix U is Nx2 in dimension.
%A stores products of elements of U that populate U'*U:

A = [U(:,1).*U(:,1), U(:,1).*U(:,2), U(:,2).*U(:,1), U(:,2).*U(:,2)];
```

The matrix A is an **Nx4** array

$\boldsymbol{A}_{11}(t_0+(k-1)\Delta t)$

$\boldsymbol{A}_{12}(t_0+(k-1)\Delta t)$

$\boldsymbol{A}_{21}(t_0+(k-1)\Delta t)$

$\boldsymbol{A}_{22}(t_0+(k-1)\Delta t)$

Module 5: Compute Test Statistic (6/10)

$$\left(\frac{N-2}{2}\right)\frac{\boldsymbol{x}_Z^{\mathrm{T}}\boldsymbol{U}\left(\boldsymbol{U}^{\mathrm{T}}\boldsymbol{U}\right)^{-1}\boldsymbol{U}^{\mathrm{T}}\boldsymbol{x}_Z}{\boldsymbol{x}_Z^{\mathrm{T}}\left(\boldsymbol{I}-\boldsymbol{U}(\boldsymbol{U}^{\mathrm{T}}\boldsymbol{U})^{-1}\boldsymbol{U}^{\mathrm{T}}\right)\boldsymbol{x}_Z}$$

$$\boldsymbol{A}=\boldsymbol{U}^{\mathrm{T}}\boldsymbol{U}=\begin{bmatrix}\boldsymbol{U}_1^T\boldsymbol{U}_1 & \boldsymbol{U}_1^T\boldsymbol{U}_2\\ \boldsymbol{U}_1^T\boldsymbol{U}_2 & \boldsymbol{U}_2^T\boldsymbol{U}_2\end{bmatrix}$$

```
%first compute U’*U: Matrix U is Nx2 in dimension.
%A stores products of elements of U that populate U’*U:

A = [U(:,1).*U(:,1), U(:,1).*U(:,2), U(:,2).*U(:,1), U(:,2).*U(:,2)];


%moving sum of U’*U: (confirmed)

At = movsum(A,[wins(1), wins(2)],'omitnan','Endpoints','shrink');
```

The matrix `At` is still an **Nx4** array

This computation replaces each value of `A` with a causal sum of `wins(1)` samples that consume data that can include noise, or noise + signal

Module 5: Compute Test Statistic (7/10)

$$\left(\frac{N-2}{2}\right)\frac{\boldsymbol{x}_Z^{\mathrm{T}}\boldsymbol{U}\left(\boldsymbol{U}^{\mathrm{T}}\boldsymbol{U}\right)^{-1}\boldsymbol{U}^{\mathrm{T}}\boldsymbol{x}_Z}{\boldsymbol{x}_Z^{\mathrm{T}}(\boldsymbol{I}-\boldsymbol{U}(\boldsymbol{U}^{\mathrm{T}}\boldsymbol{U})^{-1}\boldsymbol{U}^{\mathrm{T}})\boldsymbol{x}_Z}$$

$$\boldsymbol{A}^{-1}=\frac{1}{A_4A_1-A_2A_3}\begin{bmatrix}A_4 & -A_2\\ -A_3 & A_1\end{bmatrix}$$

```
%determinant via Cramer's rule of U'*U: (confirmed)
Cr = At(:,1).*At(:,4) - At(:,2).*At(:,3);
```

The matrix `Cr` is still an **Nx1** array

Module 5: Compute Test Statistic (8/10)

$$\left(\frac{N-2}{2}\right)\frac{\boldsymbol{x}_Z^{\mathrm{T}}\boldsymbol{U}\left(\boldsymbol{U}^{\mathrm{T}}\boldsymbol{U}\right)^{-1}\boldsymbol{U}^{\mathrm{T}}\boldsymbol{x}_Z}{\boldsymbol{x}_Z^{\mathrm{T}}(\boldsymbol{I}-\boldsymbol{U}(\boldsymbol{U}^{\mathrm{T}}\boldsymbol{U})^{-1}\boldsymbol{U}^{\mathrm{T}})\boldsymbol{x}_Z}$$

$$\boldsymbol{A}^{-1}=\frac{1}{A_4A_1-A_2A_3}\begin{bmatrix} A_4 & -A_2 \\ -A_3 & A_1 \end{bmatrix}$$

```
%determinant via Cramer's rule of U'*U: (confirmed)
Cr = At(:,1).*At(:,4) - At(:,2).*At(:,3);


%compute the inverse of U'*U, (U'*U)^(-1) (confirmed)
Ati = (1./Cr).*([At(:,4), -At(:,3), -At(:,2), At(:,1)]);
```

The matrix `Ati` is an **Nx4** array

Module 5: Compute Test Statistic (9/10)

$$\left(\frac{N-2}{2}\right)\frac{\overbrace{\boldsymbol{x}_Z^{\mathrm{T}}\boldsymbol{U}}^{1}\overbrace{\left(\boldsymbol{U}^{\mathrm{T}}\boldsymbol{U}\right)^{-1}\boldsymbol{U}^{\mathrm{T}}}^{2}\boldsymbol{x}_Z}{\boldsymbol{x}_Z^{\mathrm{T}}(\boldsymbol{I}-\boldsymbol{U}(\boldsymbol{U}^{\mathrm{T}}\boldsymbol{U})^{-1}\boldsymbol{U}^{\mathrm{T}})\boldsymbol{x}_Z}$$

The matrix At is still an **Nx4** array

```
%compute the product U'*x (confirmed):
```
1
```
Atx = movsum([U(:,1).*x(:), U(:,2).*x(:)],[wins(1), wins(2)],...
    'omitnan','Endpoints','shrink');
```

```
%compute the product ((U’*U).^1)*U'*x: (confirmed)
```
2
```
Pest = [Atx(:,1).*Ati(:,1) + Atx(:,2).*Ati(:,3), Atx(:,1).*Ati(:,2)
       + Atx(:,2).*Ati(:,4)];
```

```
%compute the product x’*U*((U’*U).^1)*U'*x: (confirmed) as the norm
of the projection onto the subspace spanned by the columns of U.
Proj = Pest(:,1).*Atx(:,1) + Pest(:,2).*Atx(:,2);
```

Last matrix is an **Nx1** array

Module 5: Compute Test Statistic (10/10)

$$\left(\frac{N-2}{2}\right)\frac{\boldsymbol{x}_Z^{\mathrm{T}}\boldsymbol{U}\left(\boldsymbol{U}^{\mathrm{T}}\boldsymbol{U}\right)^{-1}\boldsymbol{U}^{\mathrm{T}}\boldsymbol{x}_Z}{\boldsymbol{x}_Z^{\mathrm{T}}(\boldsymbol{I}-\boldsymbol{U}(\boldsymbol{U}^{\mathrm{T}}\boldsymbol{U})^{-1}\boldsymbol{U}^{\mathrm{T}})\boldsymbol{x}_Z}$$

Previous slides summarized numerator computation

Similar computation outputs denominator

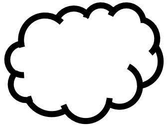
```
%compute the product x'*(I - U*((U'*U).^1)*U')*x: (confirmed) as the
%norm of the projection onto the subspace spanned by the columns of U.

Perp = movsum(x.*x,[wins(1), wins(2)],'omitnan','Endpoints',...
       'shrink') - Proj;
```

Summary points:
Vectorize arithmetic to store matrix elements as columns, time indices as rows, use causal windows to sum backward, and compute matrix inverses over sliding windows.
Of course, function `movsum.m` helps.

Module 6:
Estimate Parameters & Thresholds

Status Update: What we've Done Already

Single modality detection: terminology, concepts



$\mathcal{H}_0$: $\boldsymbol{x} =$ noise + background

$\mathcal{H}_1$: $\boldsymbol{x} =$ noise + background + signal

$$\frac{\widehat{\mathrm{PDF}}(\boldsymbol{x}|\mathcal{H}_1)}{\widehat{\mathrm{PDF}}(\boldsymbol{x}|\mathcal{H}_0)} = s(\boldsymbol{x}) \underset{\mathcal{H}_0}{\overset{\mathcal{H}_1}{\gtrless}} t \qquad \mathcal{H}_1 \rightarrow u = 1, \quad \mathcal{H}_0 \rightarrow u = 0$$

Source | **Data Models** | **GLR** | **Detector** | **Decision**

Density processing: estimate distributional parameters that shape density and histograms

Data pre-processing: detrend, taper and bandpass filter (possibly in a filter bank)

Status Update: What we've Done Already

Single modality detection: terminology, concepts

$\mathcal{H}_0$: $\boldsymbol{x} =$ noise + background

$\mathcal{H}_1$: $\boldsymbol{x} =$ noise + background + signal

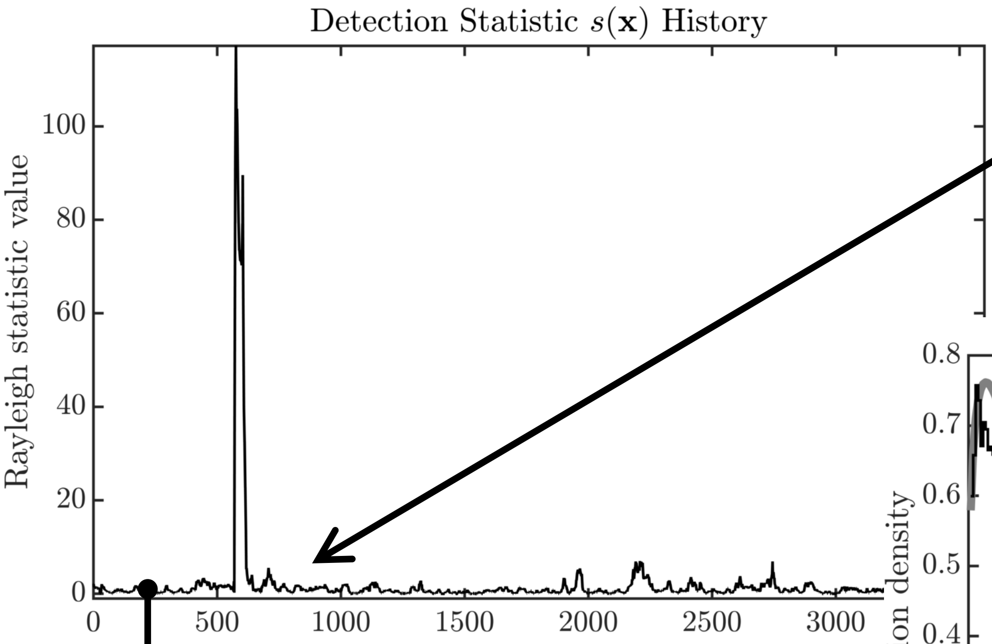
$$\frac{\widehat{\mathrm{PDF}}(\boldsymbol{x}|\mathcal{H}_1)}{\widehat{\mathrm{PDF}}(\boldsymbol{x}|\mathcal{H}_0)} = s(\boldsymbol{x}) \underset{\mathcal{H}_0}{\overset{\mathcal{H}_1}{\gtrless}} t \qquad \begin{matrix} \mathcal{H}_1 \rightarrow u = 1 \\ \mathcal{H}_0 \rightarrow u = 0 \end{matrix}$$

Source | **Data Models** | **GLR** | **Detector** | **Decision**

When the detection statistic exceeds threshold t, the detector declares the presence of a target signal, and that hypothesis $\mathcal{H}_1$ is true.

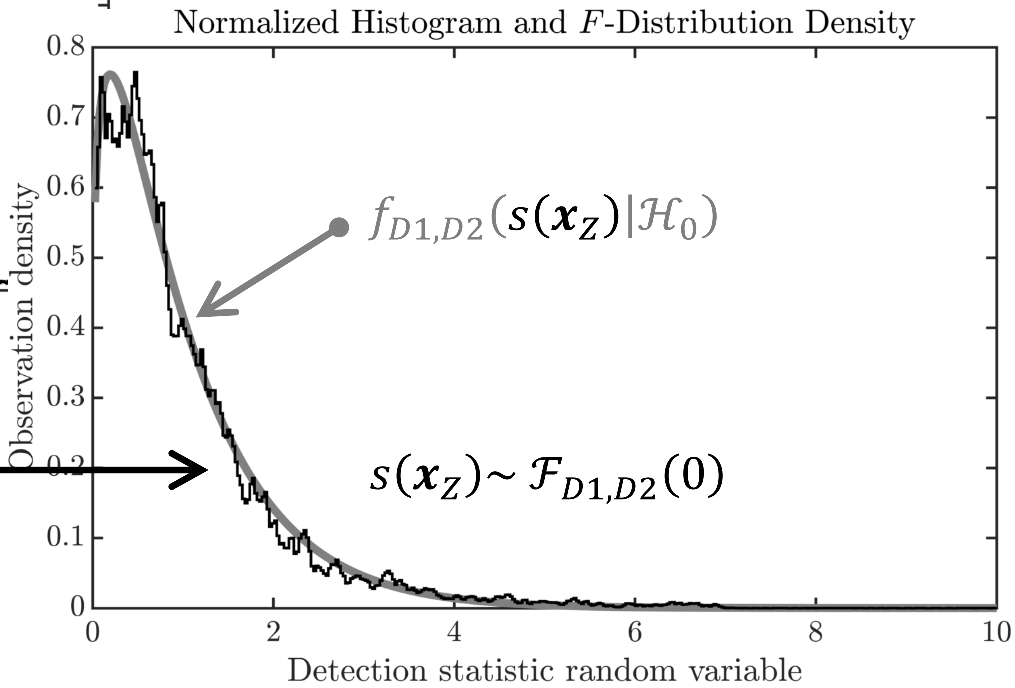
Data pre-processing: detrend, taper and bandpass filter (possibly in a filter bank)

Module 6: Estimate Parameters and Thresholds (1/11)



$$s(\boldsymbol{x}_Z) = \left(\frac{N-2}{2}\right) \frac{\boldsymbol{x}_Z^{\mathrm{T}} \boldsymbol{U}\left(\boldsymbol{U}^{\mathrm{T}} \boldsymbol{U}\right)^{-1} \boldsymbol{U}^{\mathrm{T}} \boldsymbol{x}_Z}{\boldsymbol{x}_Z^{\mathrm{T}}(\boldsymbol{I} - \boldsymbol{U}(\boldsymbol{U}^{\mathrm{T}} \boldsymbol{U})^{-1} \boldsymbol{U}^{\mathrm{T}}) \boldsymbol{x}_Z}$$

Histogram$[s(\boldsymbol{x}_Z)]$

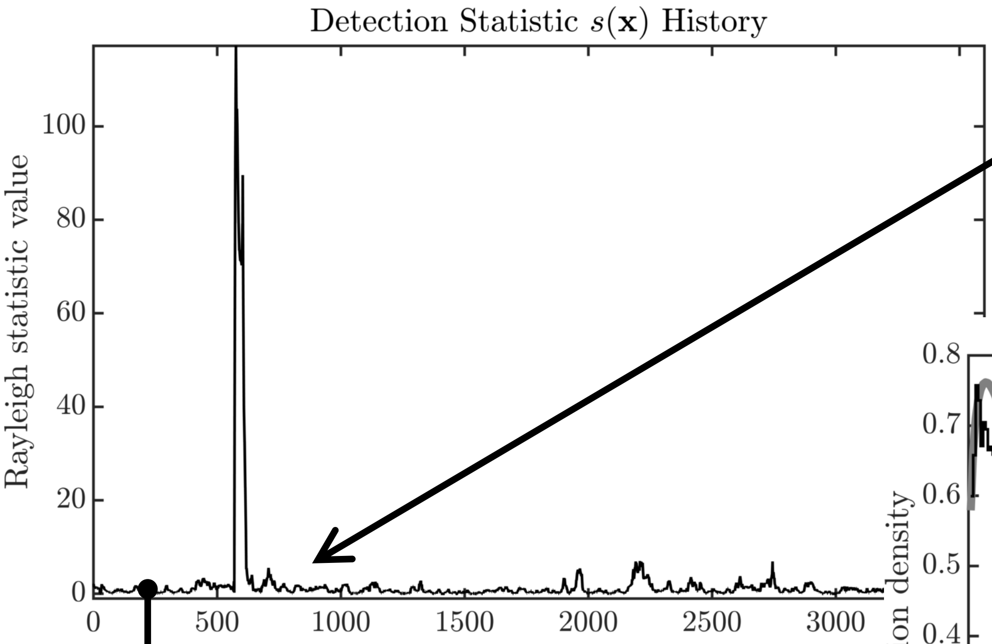


$$f_{D1,D2}(s(\boldsymbol{x}_Z)|\mathcal{H}_0)$$

$$s(\boldsymbol{x}_Z) \sim \mathcal{F}_{D1,D2}(0)$$

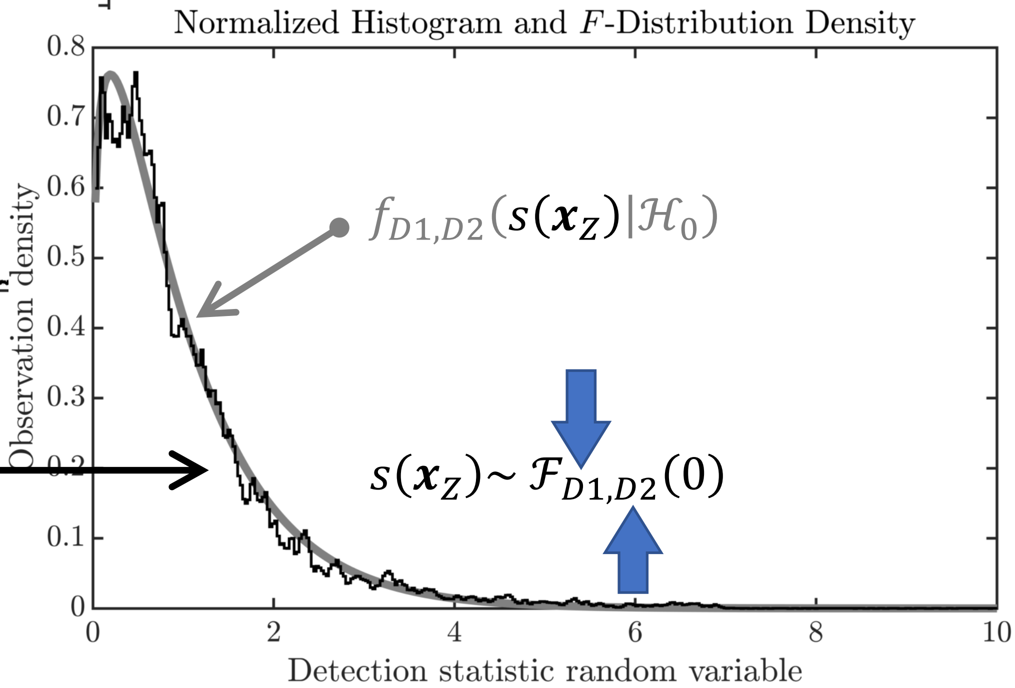
Los Alamos NATIONAL LABORATORY

Module 6: Estimate Parameters and Thresholds (2/11)



$$s(\boldsymbol{x}_Z) = \left(\frac{N-2}{2}\right) \frac{\boldsymbol{x}_Z^{\mathrm{T}} \boldsymbol{U}\left(\boldsymbol{U}^{\mathrm{T}} \boldsymbol{U}\right)^{-1} \boldsymbol{U}^{\mathrm{T}} \boldsymbol{x}_Z}{\boldsymbol{x}_Z^{\mathrm{T}}\left(\boldsymbol{I} - \boldsymbol{U}\left(\boldsymbol{U}^{\mathrm{T}} \boldsymbol{U}\right)^{-1} \boldsymbol{U}^{\mathrm{T}}\right) \boldsymbol{x}_Z}$$

Histogram$[s(\boldsymbol{x}_Z)]$

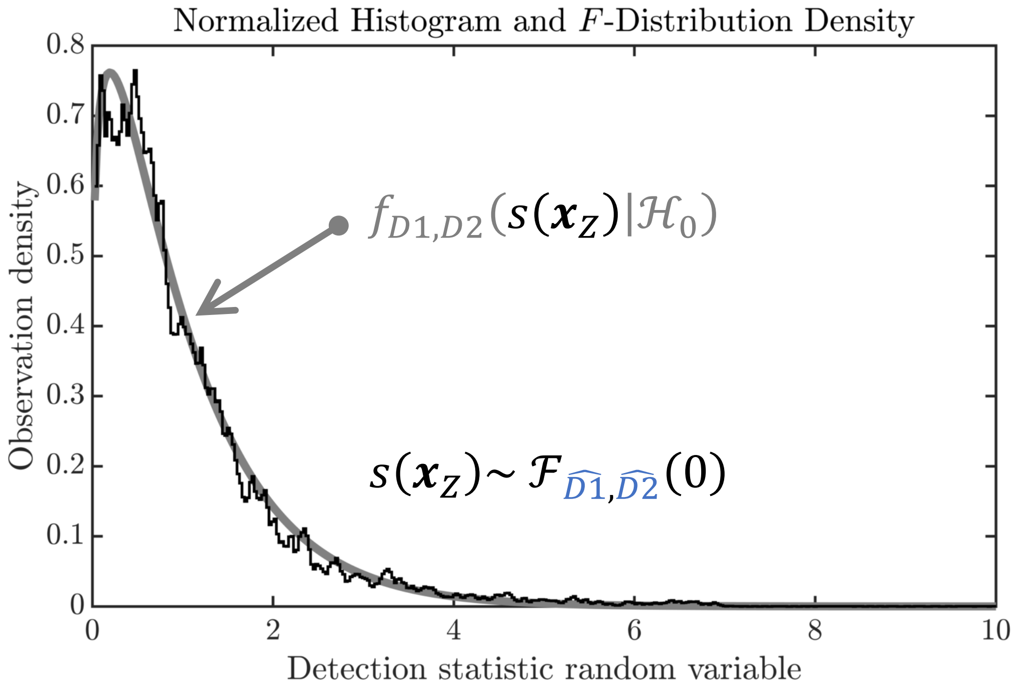


$$s(\boldsymbol{x}_Z) \sim \mathcal{F}_{D1,D2}(0)$$

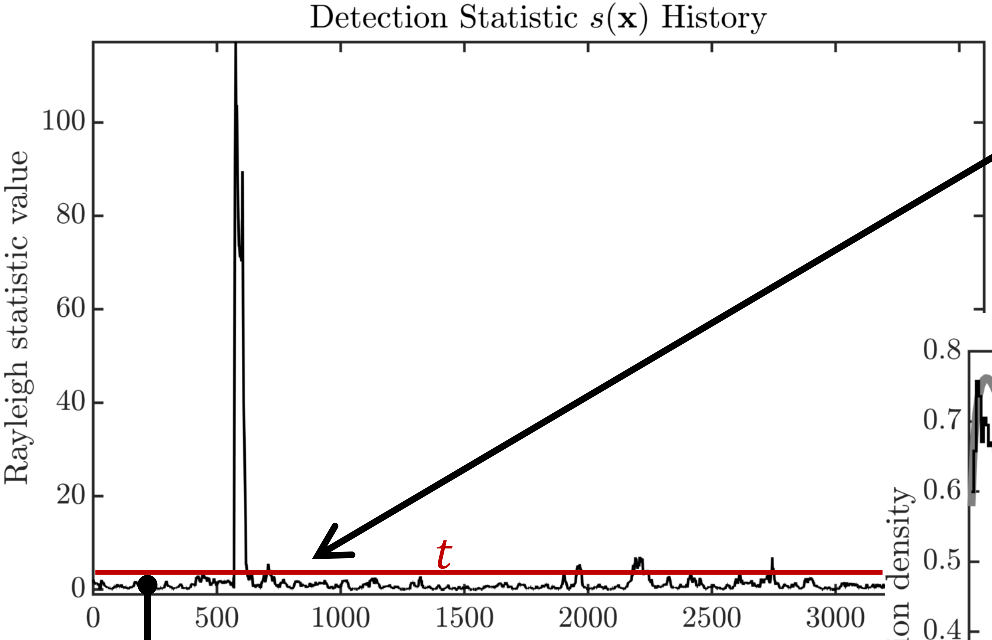
Module 6: Estimate Parameters and Thresholds (3/11)

$$\widehat{D1}, \widehat{D2} = \underset{D1,D2}{\operatorname{argmin}} \left\| \operatorname{Hist}\big(s(\boldsymbol{x}_Z)\big)_{2.5}^{95} - f_{D1,D2}\big(s(\boldsymbol{x}_Z)|\mathcal{H}_0\big) \right\|$$

In words: estimate the degree of freedom parameters to minimize mismatch between the normalized histogram and a theoretical central *F* density function.

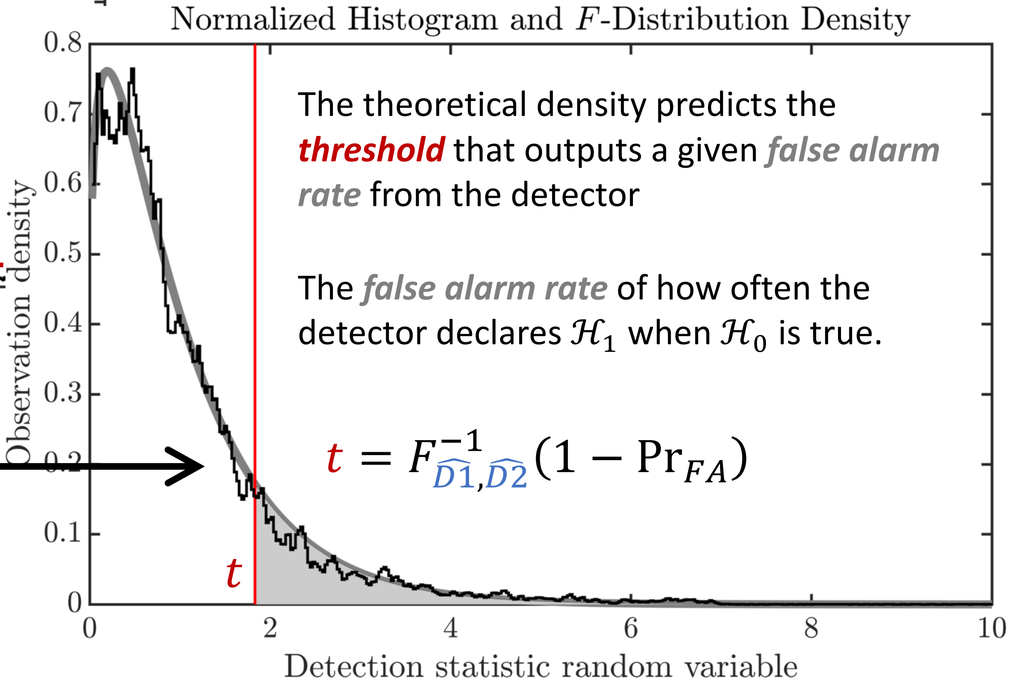


Module 6: Estimate Parameters and Thresholds (4/11)



$$s(\boldsymbol{x}_Z) = \left(\frac{N-2}{2}\right) \frac{\boldsymbol{x}_Z^{\mathrm{T}} \boldsymbol{U}\left(\boldsymbol{U}^{\mathrm{T}}\boldsymbol{U}\right)^{-1} \boldsymbol{U}^{\mathrm{T}} \boldsymbol{x}_Z}{\boldsymbol{x}_Z^{\mathrm{T}}(\boldsymbol{I} - \boldsymbol{U}(\boldsymbol{U}^{\mathrm{T}}\boldsymbol{U})^{-1}\boldsymbol{U}^{\mathrm{T}})\boldsymbol{x}_Z}$$

Histogram$[s(\boldsymbol{x}_Z)]$

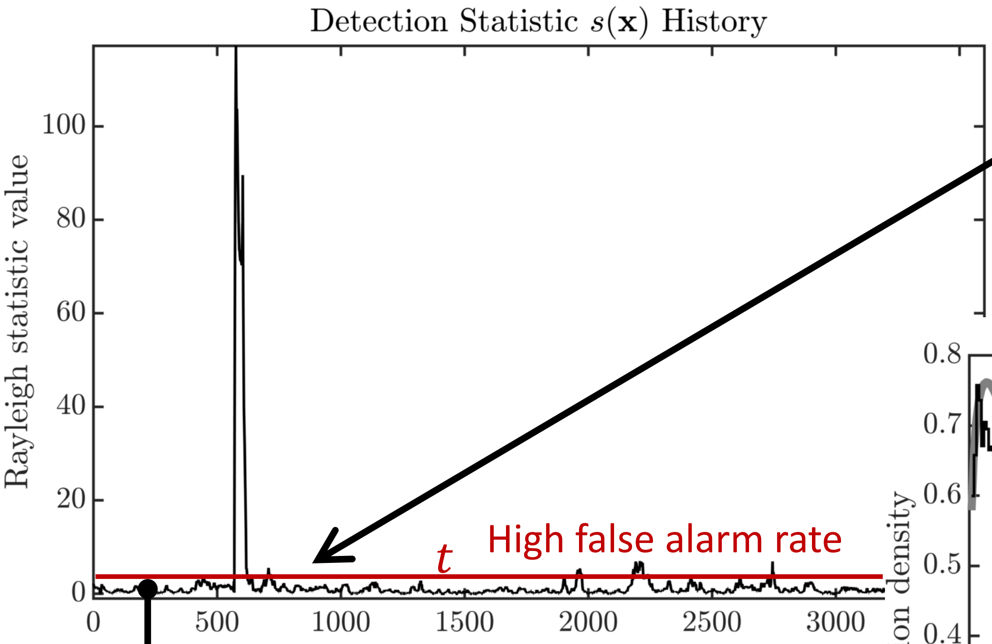


The theoretical density predicts the ***threshold*** that outputs a given ***false alarm rate*** from the detector

The ***false alarm rate*** of how often the detector declares $\mathcal{H}_1$ when $\mathcal{H}_0$ is true.

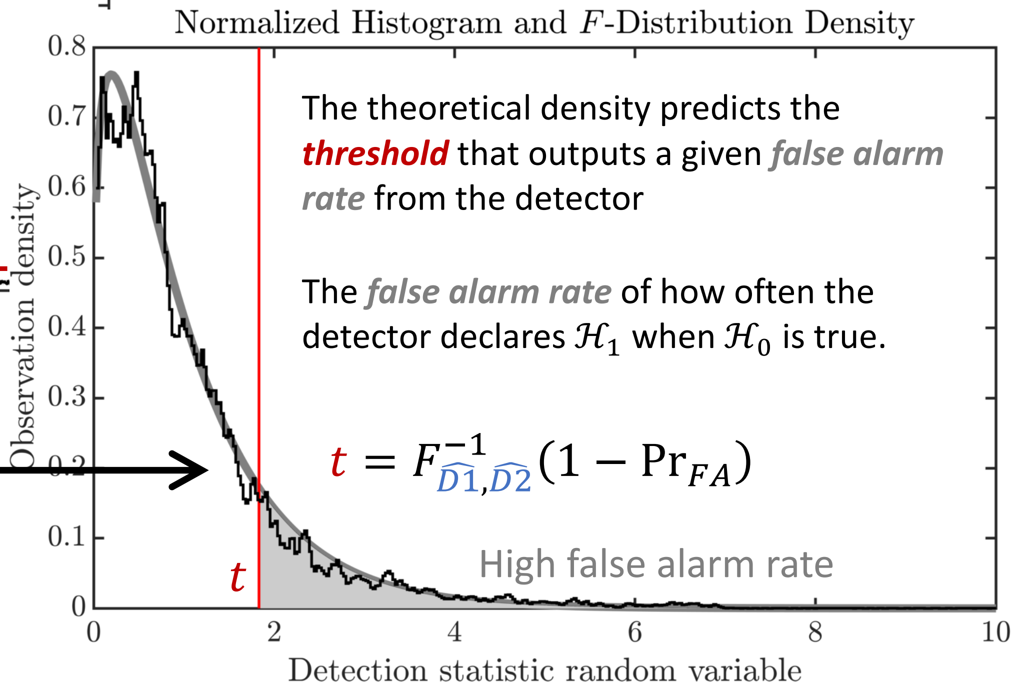
$$t = F^{-1}_{\widehat{D1},\widehat{D2}}(1 - \Pr_{FA})$$

Module 6: Estimate Parameters and Thresholds (5/11)



$$s(\boldsymbol{x}_Z) = \left(\frac{N-2}{2}\right) \frac{\boldsymbol{x}_Z^{\mathrm{T}} \boldsymbol{U}\left(\boldsymbol{U}^{\mathrm{T}}\boldsymbol{U}\right)^{-1}\boldsymbol{U}^{\mathrm{T}}\boldsymbol{x}_Z}{\boldsymbol{x}_Z^{\mathrm{T}}\left(\boldsymbol{I} - \boldsymbol{U}(\boldsymbol{U}^{\mathrm{T}}\boldsymbol{U})^{-1}\boldsymbol{U}^{\mathrm{T}}\right)\boldsymbol{x}_Z}$$

Histogram$[s(\boldsymbol{x}_Z)]$



The theoretical density predicts the ***threshold*** that outputs a given ***false alarm rate*** from the detector

The ***false alarm rate*** of how often the detector declares $\mathcal{H}_1$ when $\mathcal{H}_0$ is true.

$$t = F^{-1}_{\widehat{D1},\widehat{D2}}(1 - \mathrm{Pr}_{FA})$$

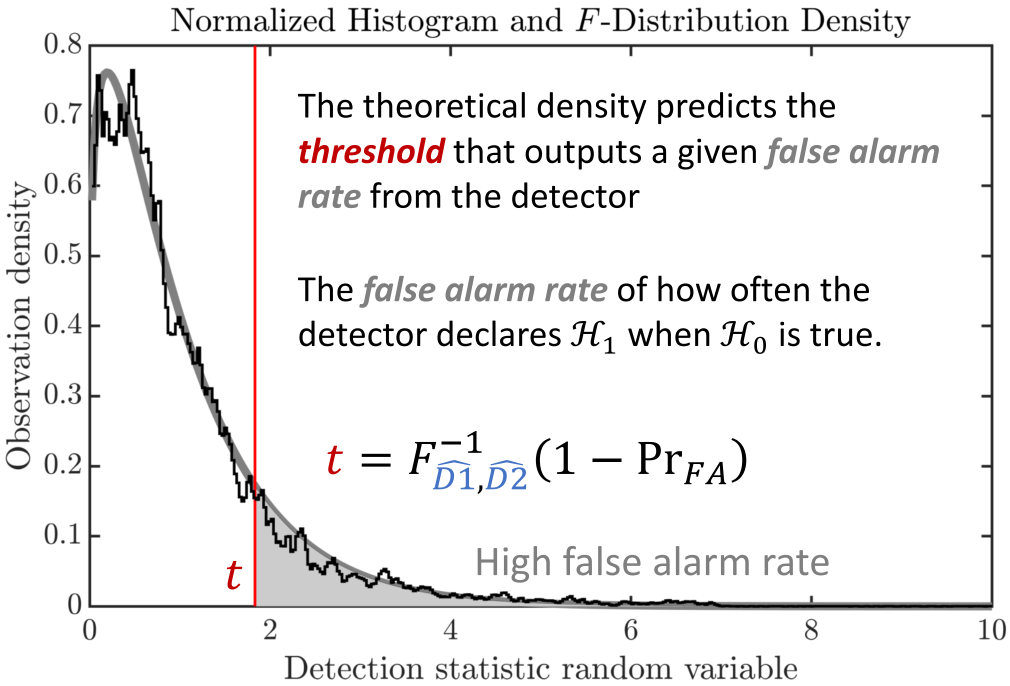
Module 6: Estimate Parameters and Thresholds (6/11)

CFAR thresholds are not *p*-values

- A threshold is a fixed value you select. You invert for it. You do not observe it.
- A p-value is an observation. If the t present in the plot was $t = s(x_Z^t)$, it could be used to compute a **p-value**.
- The p-values are computed from the **null distribution** and can provide equivalent information to the detection statistic.

Compute p-values from the density, or the cumulative distribution.

$$s(\boldsymbol{x}_Z) = \left(\frac{N-2}{2}\right) \frac{\boldsymbol{x}_Z^{\mathrm{T}} \boldsymbol{U}\left(\boldsymbol{U}^{\mathrm{T}}\boldsymbol{U}\right)^{-1}\boldsymbol{U}^{\mathrm{T}}\boldsymbol{x}_Z}{\boldsymbol{x}_Z^{\mathrm{T}}(\boldsymbol{I} - \boldsymbol{U}(\boldsymbol{U}^{\mathrm{T}}\boldsymbol{U})^{-1}\boldsymbol{U}^{\mathrm{T}})\boldsymbol{x}_Z}$$

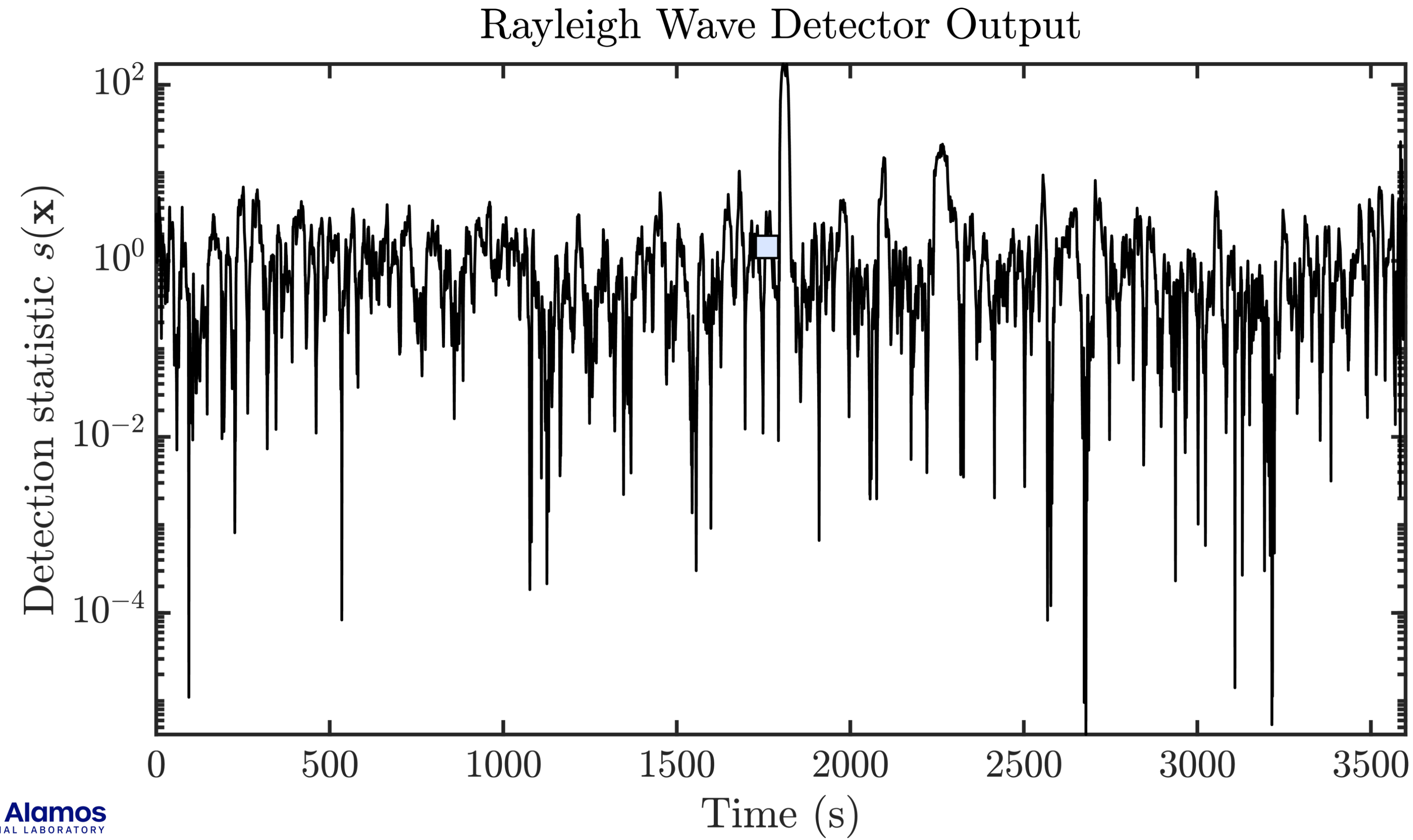


The theoretical density predicts the ***threshold*** that outputs a given ***false alarm rate*** from the detector

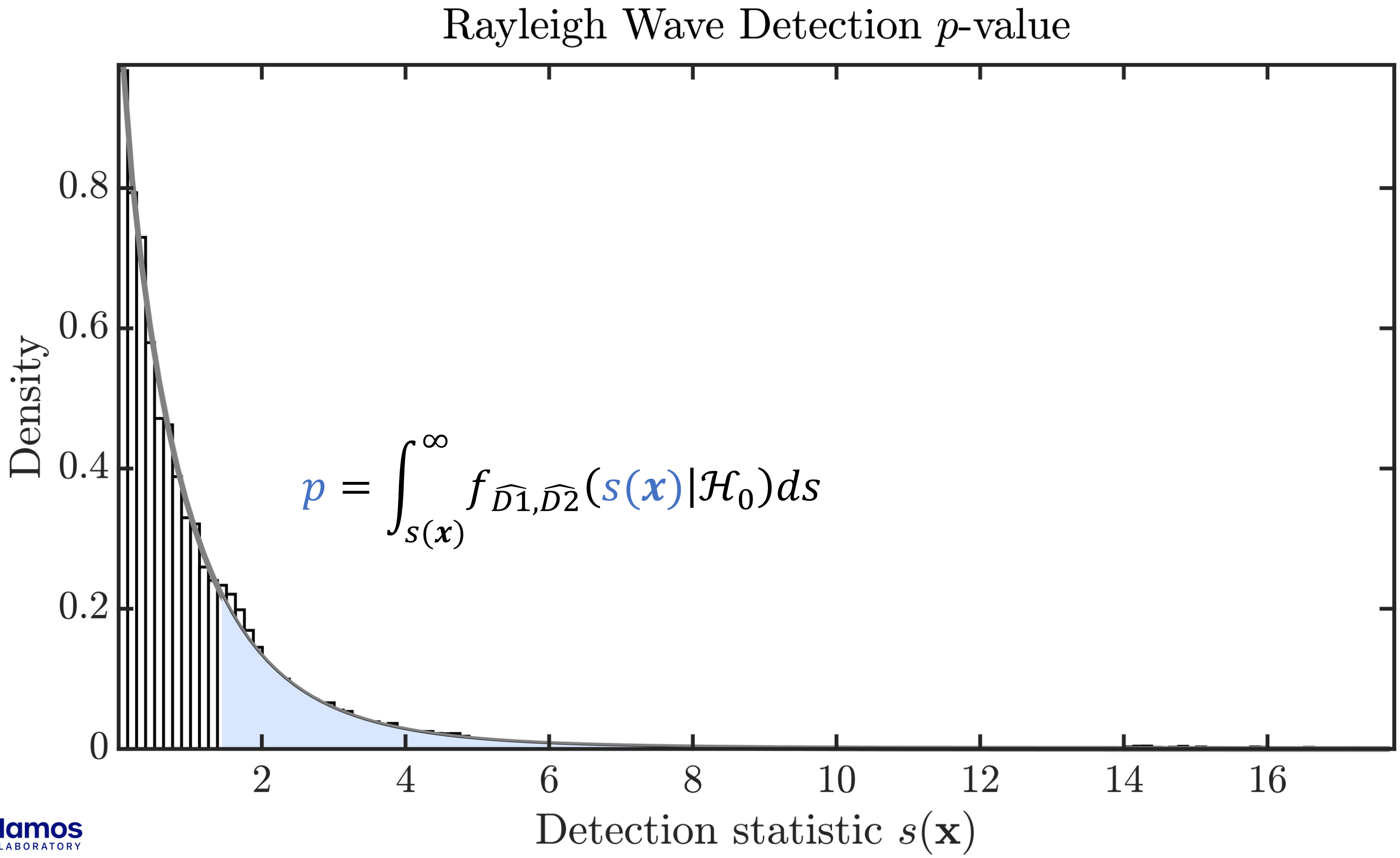
The ***false alarm rate*** of how often the detector declares $\mathcal{H}_1$ when $\mathcal{H}_0$ is true.

$$t = F^{-1}_{\widehat{D1},\widehat{D2}}(1 - \mathrm{Pr}_{FA})$$

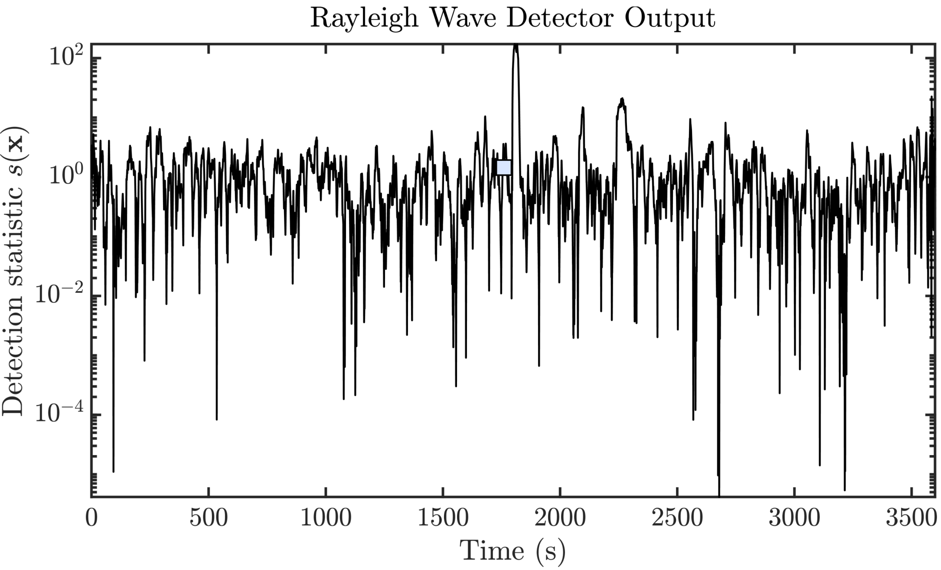
Module 6: Estimate Parameters and Thresholds (7/11)

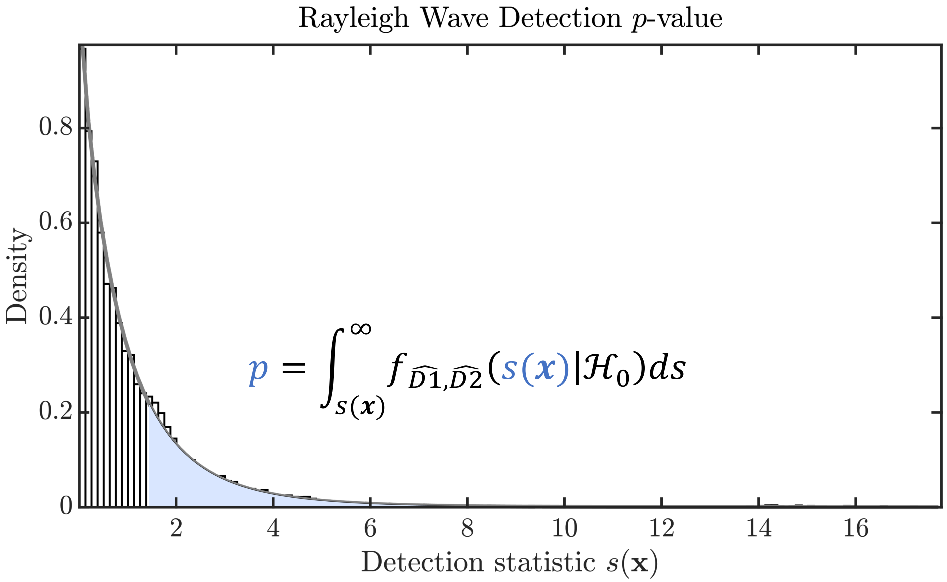


Module 6: Estimate Parameters and Thresholds (8/11)



Module 6: Estimate Parameters and Thresholds (9/11)

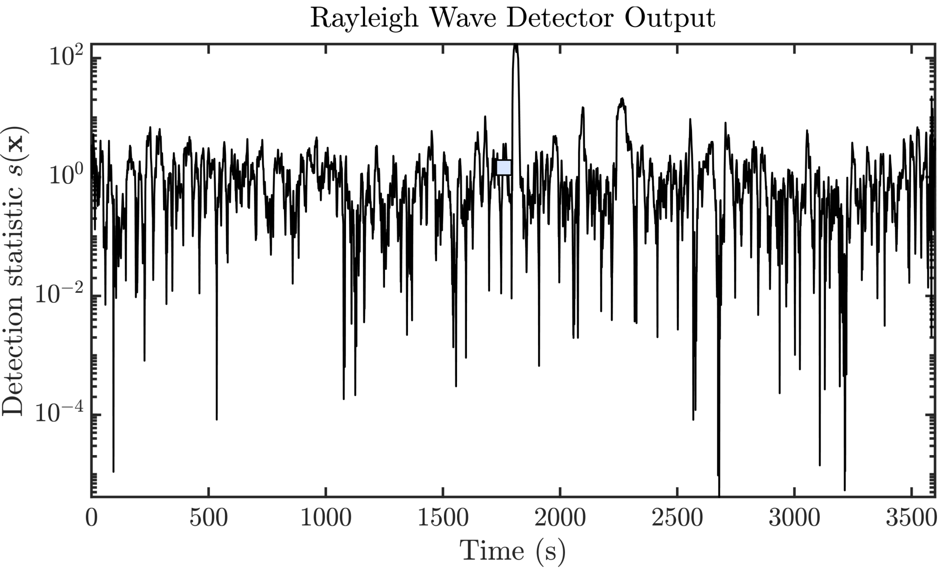


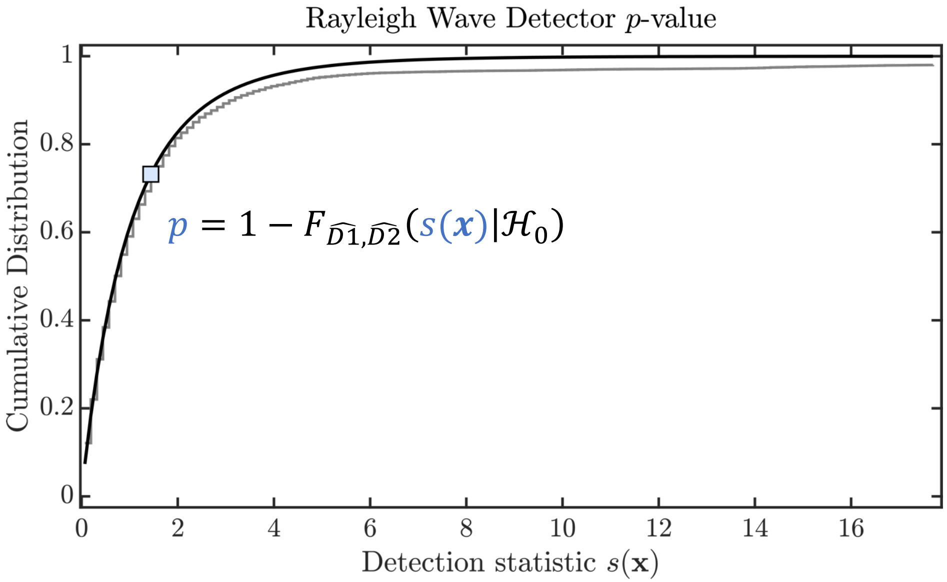


Select a particular detection statistic value as a sample to compute a p-value

Compute the area under the null density **fit** to estimate the p-value, or just use the empirical data (if you have enough).

Module 6: Estimate Parameters and Thresholds (10/11)

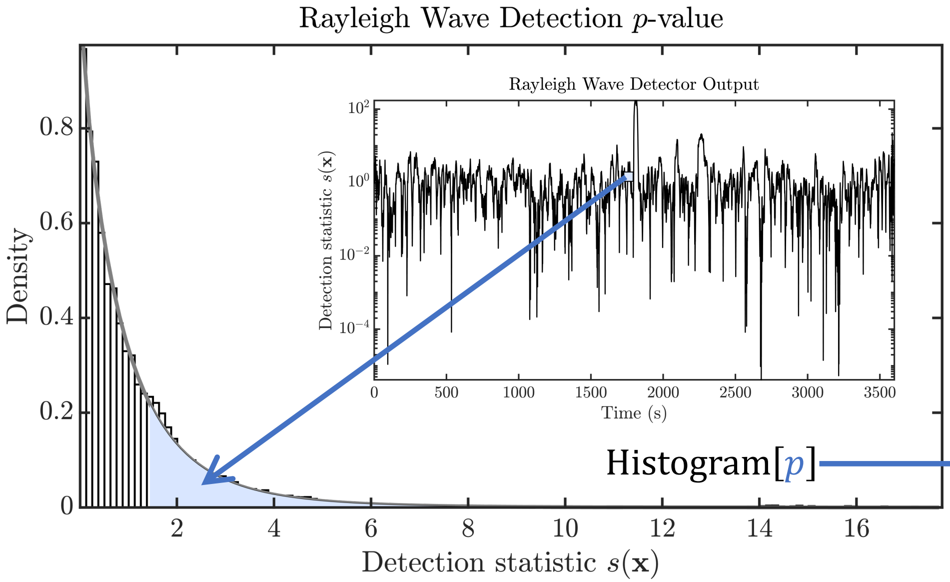


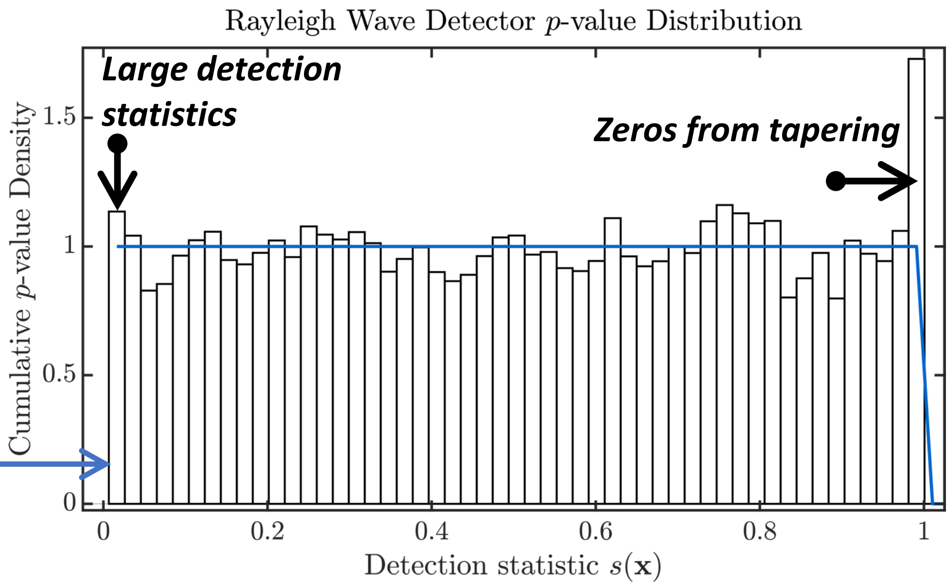


Select a particular detection statistic value as a sample to compute a p-value

Equivalently, use the cumulative distribution to estimate the p-value, or just use the empirical data (**again,** if you have enough).

Module 6: Estimate Parameters and Thresholds (11/11)





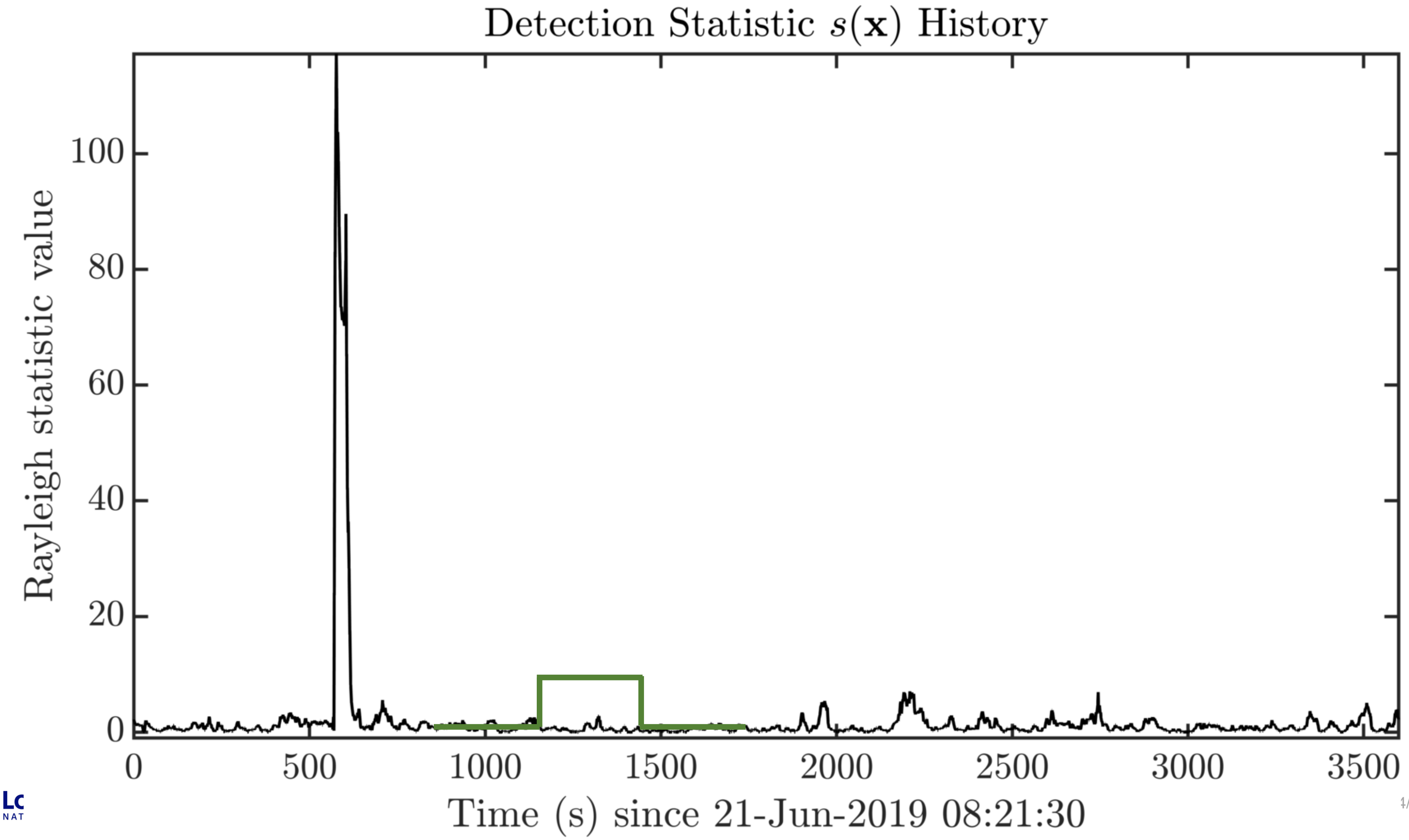
Bin all the p-values output from the detection statistic time-series

The p-value density is uniform when the observations are from the null. When they're *not*, the density shows a peak near **one**.

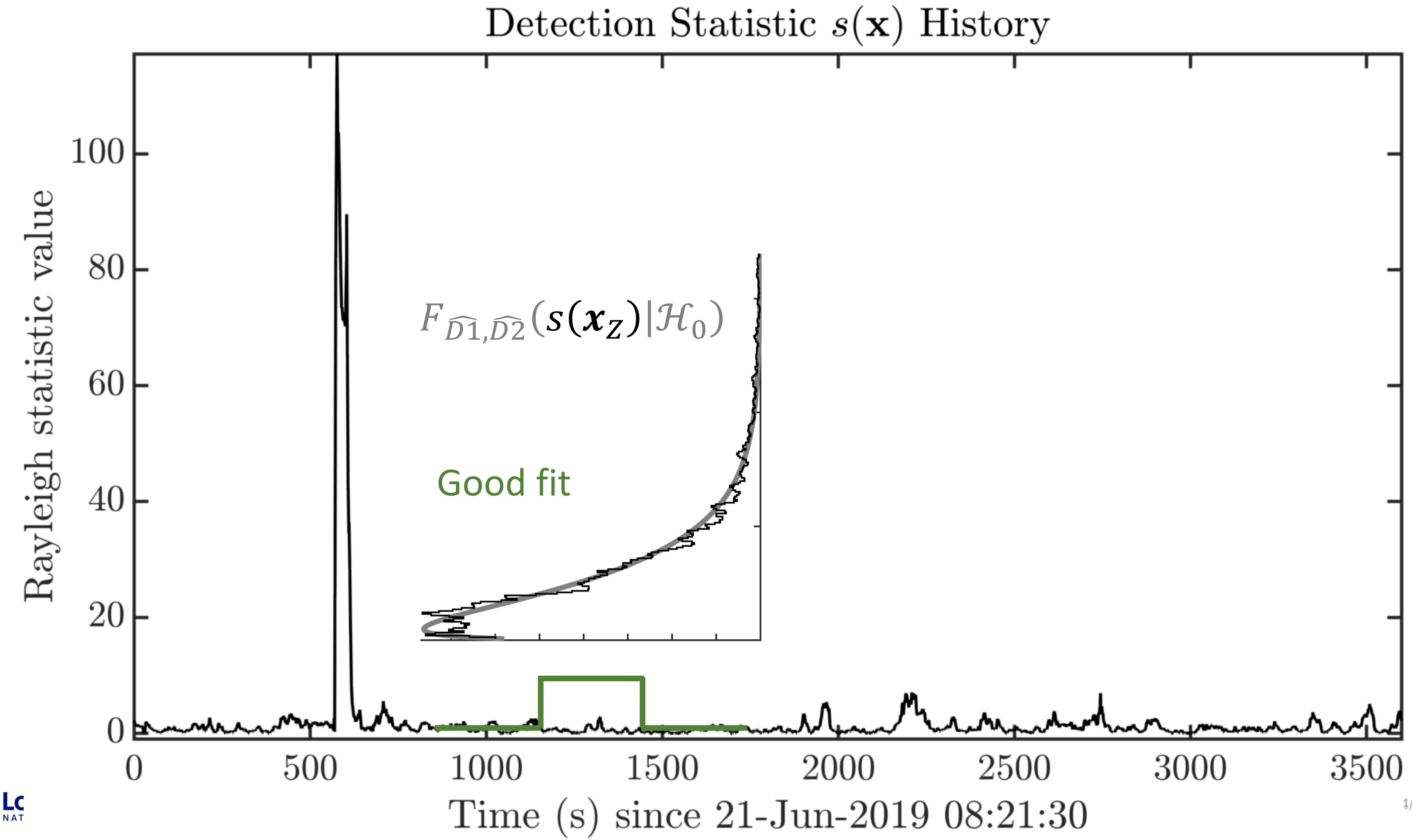
Los Alamos
NATIONAL LABORATORY

Module 7:
Adaptive Bayesian Thresholds

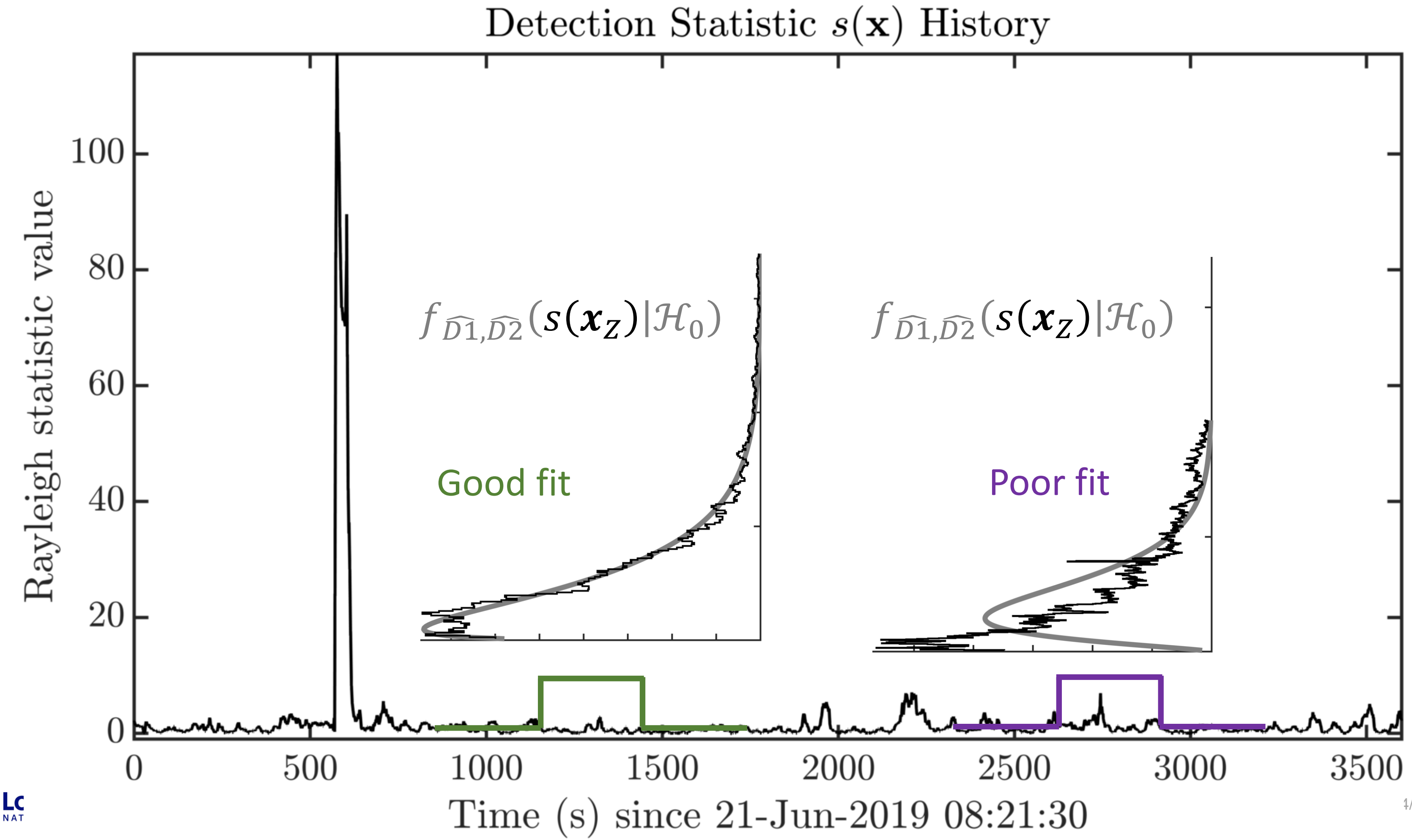
Module 7: Adaptive Bayesian Thresholds (1/11)



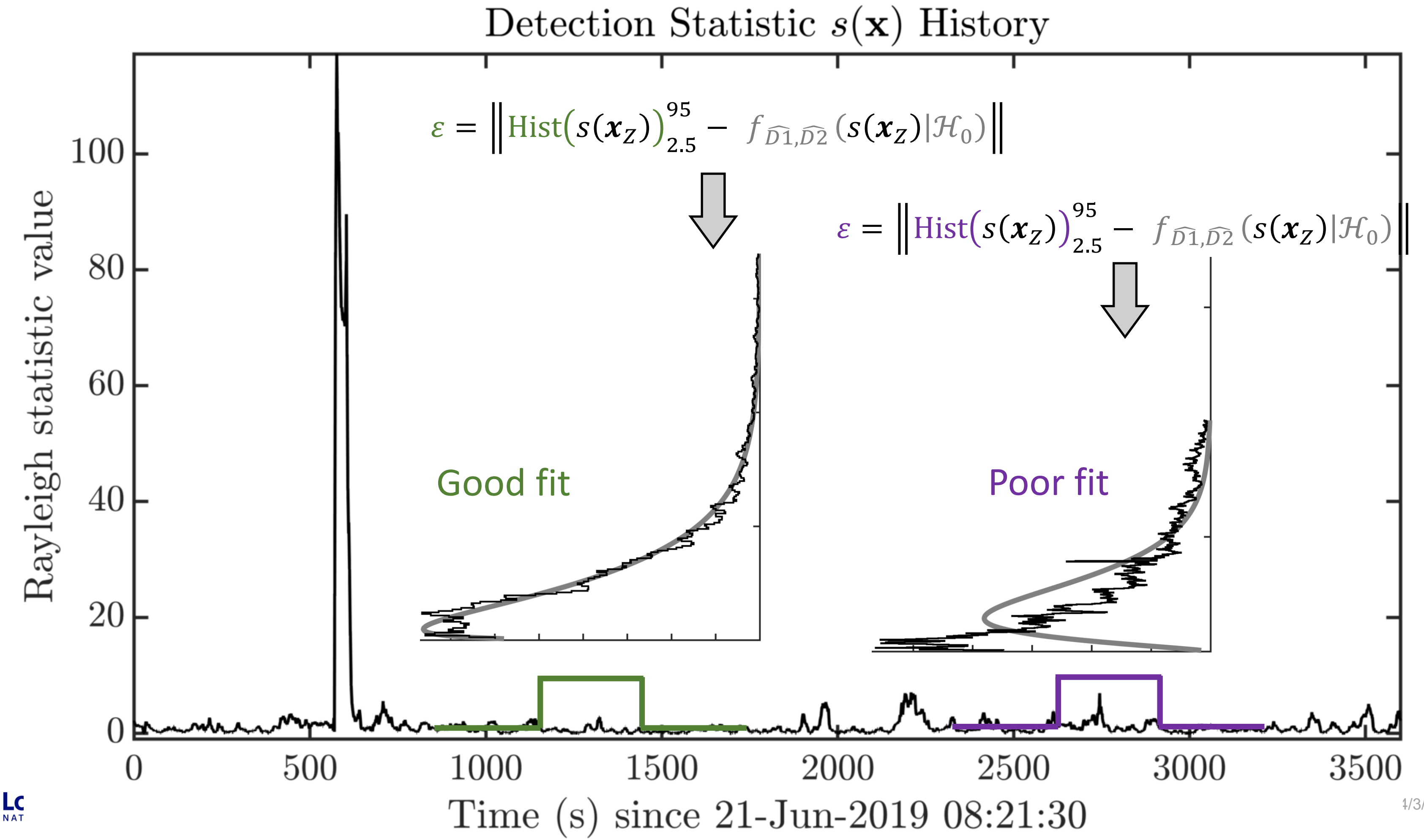
Module 7: Adaptive Bayesian Thresholds (2/11)



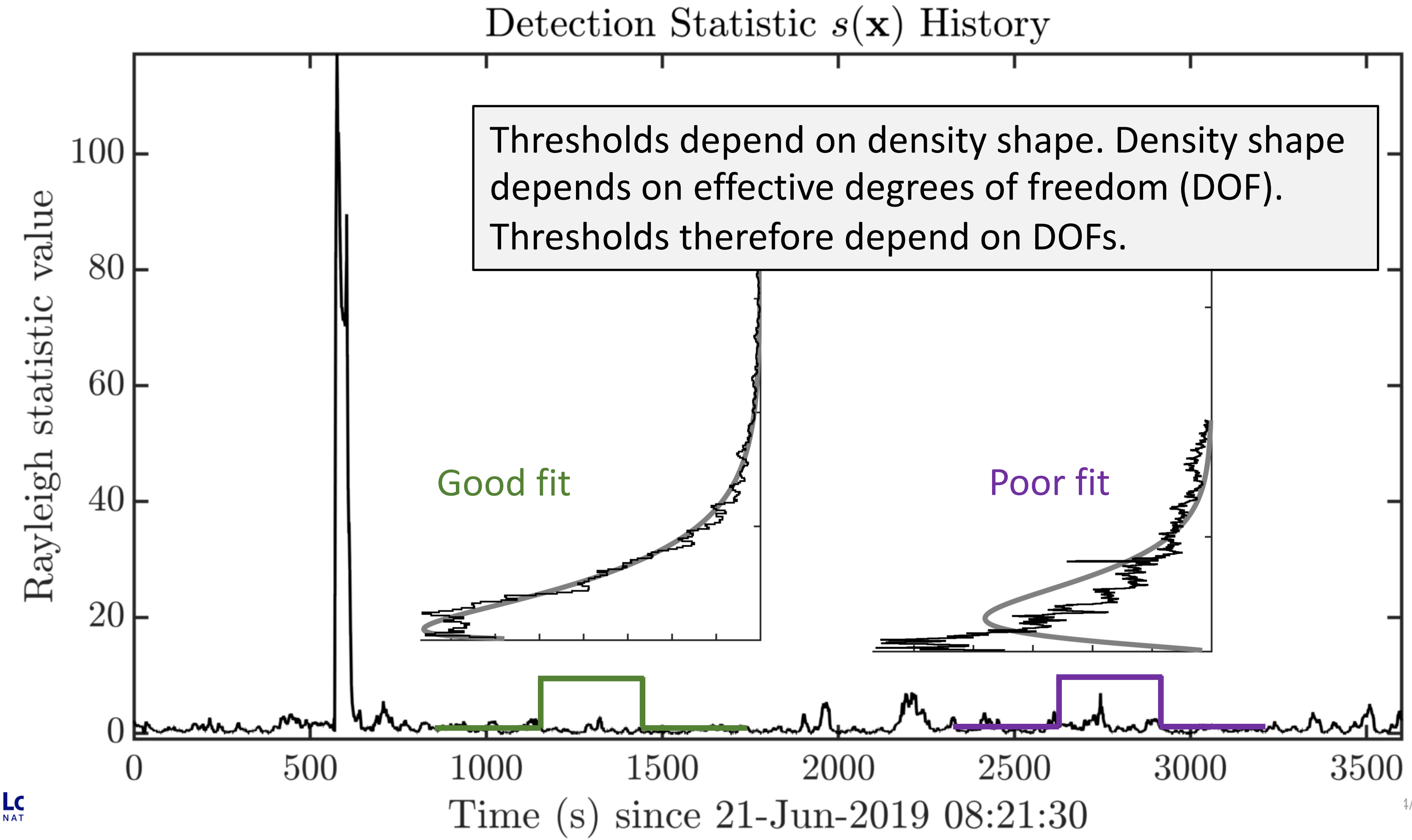
Module 7: Adaptive Bayesian Thresholds (3/11)



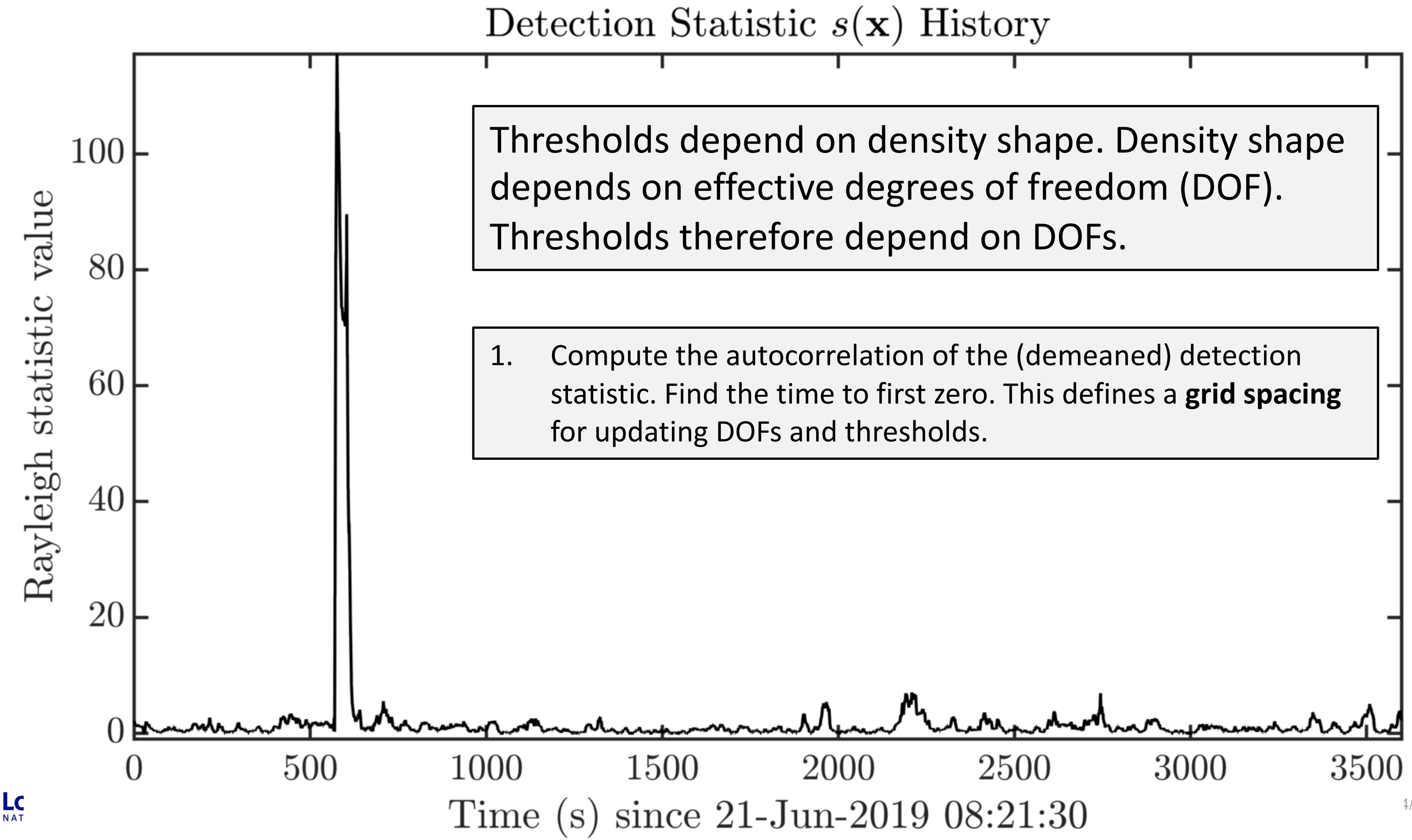
Module 7: Adaptive Bayesian Thresholds (1/11)



Module 7: Adaptive Bayesian Thresholds (5/11)



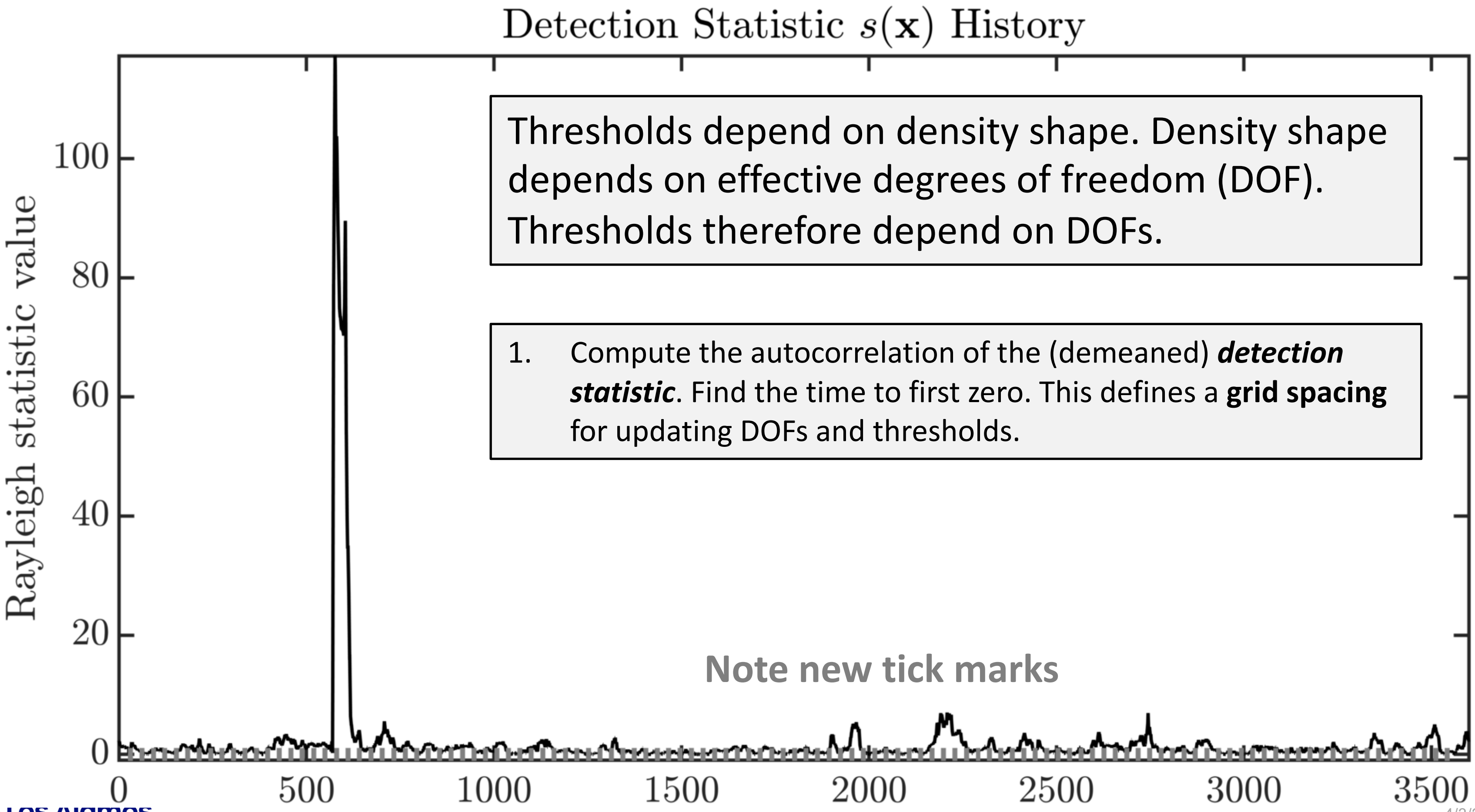
Module 6: Adaptive Bayesian Thresholds (6/11)



Thresholds depend on density shape. Density shape depends on effective degrees of freedom (DOF). Thresholds therefore depend on DOFs.

1. Compute the autocorrelation of the (demeaned) detection statistic. Find the time to first zero. This defines a **grid spacing** for updating DOFs and thresholds.

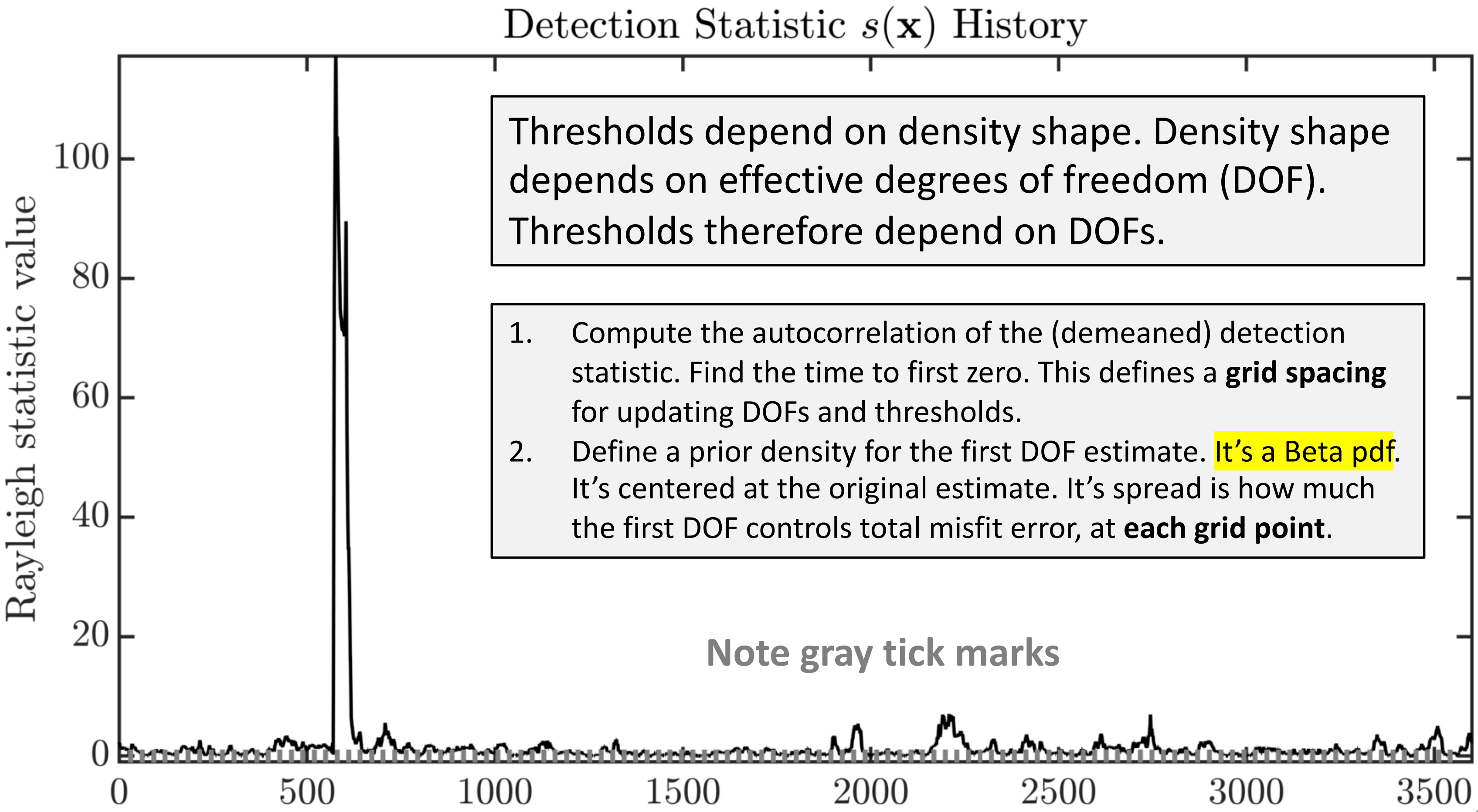
Module 6: Adaptive Bayesian Thresholds (7/11)



Thresholds depend on density shape. Density shape depends on effective degrees of freedom (DOF). Thresholds therefore depend on DOFs.

1. Compute the autocorrelation of the (demeaned) ***detection statistic***. Find the time to first zero. This defines a **grid spacing** for updating DOFs and thresholds.

Module 6: Adaptive Bayesian Thresholds (8/11)

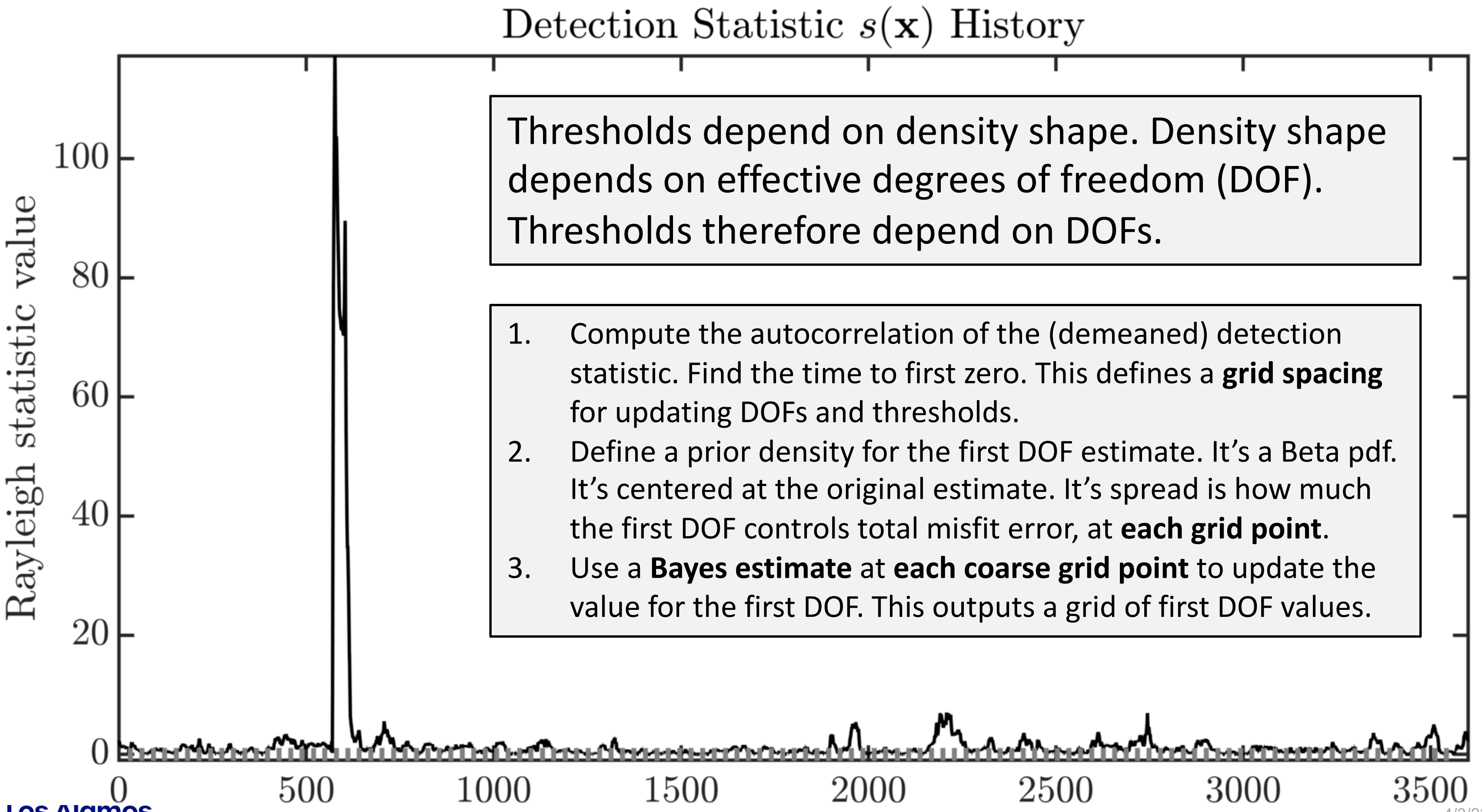


Thresholds depend on density shape. Density shape depends on effective degrees of freedom (DOF). Thresholds therefore depend on DOFs.

1. Compute the autocorrelation of the (demeaned) detection statistic. Find the time to first zero. This defines a **grid spacing** for updating DOFs and thresholds.
2. Define a prior density for the first DOF estimate. It's a Beta pdf. It's centered at the original estimate. It's spread is how much the first DOF controls total misfit error, at **each grid point**.

Note gray tick marks

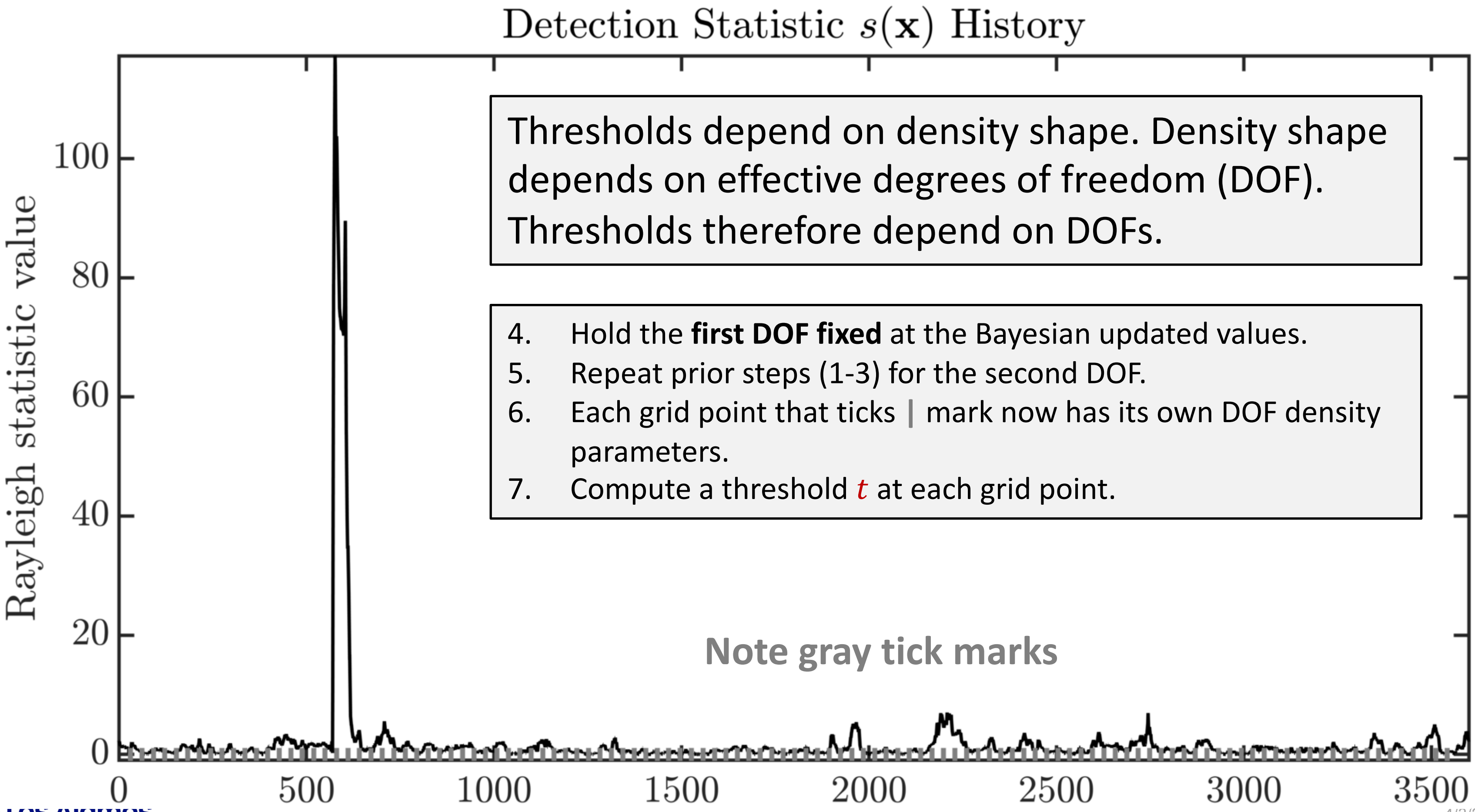
Module 6: Adaptive Bayesian Thresholds (9/11)



Thresholds depend on density shape. Density shape depends on effective degrees of freedom (DOF). Thresholds therefore depend on DOFs.

1. Compute the autocorrelation of the (demeaned) detection statistic. Find the time to first zero. This defines a **grid spacing** for updating DOFs and thresholds.
2. Define a prior density for the first DOF estimate. It’s a Beta pdf. It’s centered at the original estimate. It’s spread is how much the first DOF controls total misfit error, at **each grid point**.
3. Use a **Bayes estimate** at **each coarse grid point** to update the value for the first DOF. This outputs a grid of first DOF values.

Module 6: Adaptive Bayesian Thresholds (10/11)

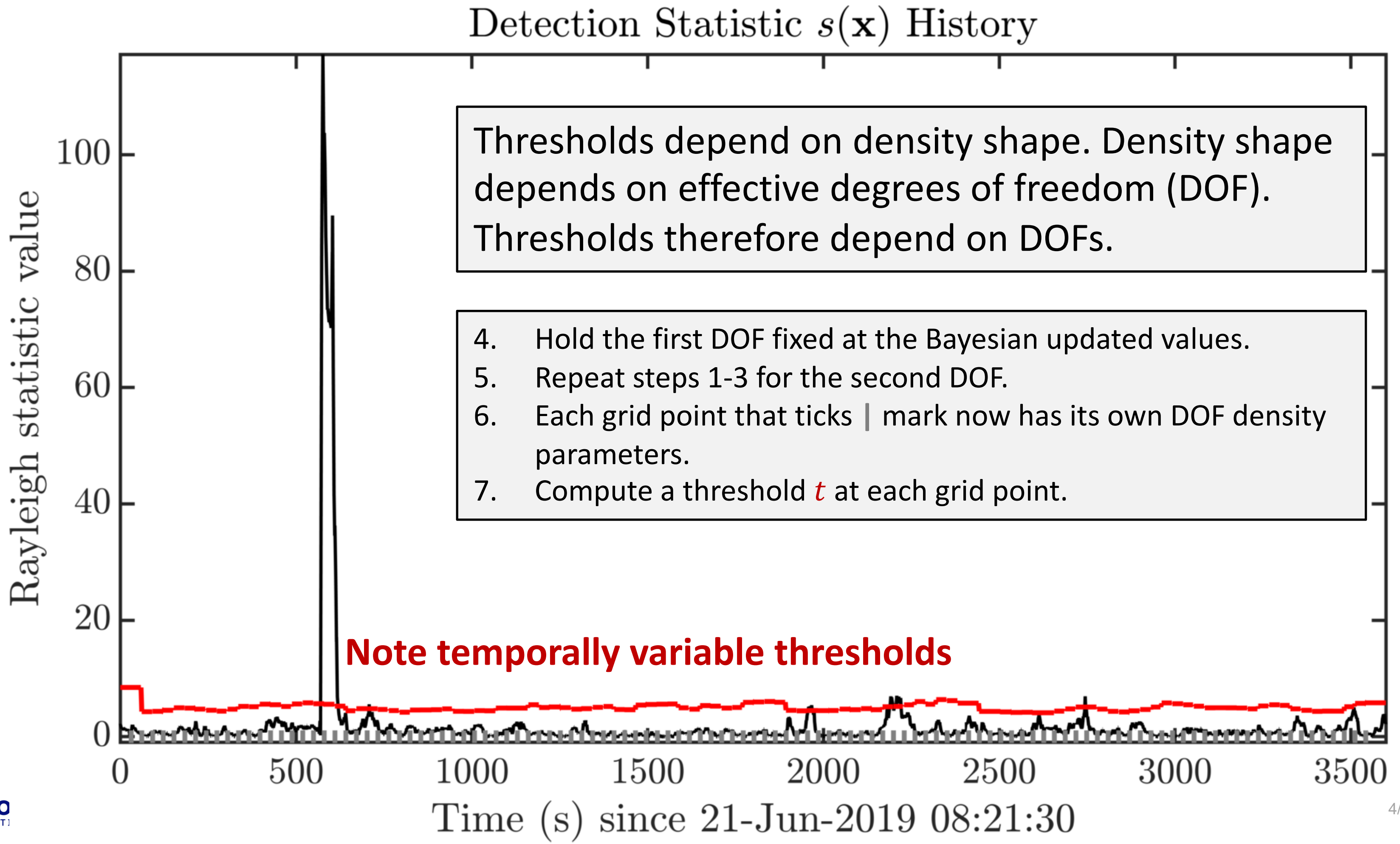


Thresholds depend on density shape. Density shape depends on effective degrees of freedom (DOF). Thresholds therefore depend on DOFs.

4. Hold the **first DOF fixed** at the Bayesian updated values.
5. Repeat prior steps (1-3) for the second DOF.
6. Each grid point that ticks | mark now has its own DOF density parameters.
7. Compute a threshold t at each grid point.

Note gray tick marks

Module 6: Adaptive Bayesian Thresholds (11/11)

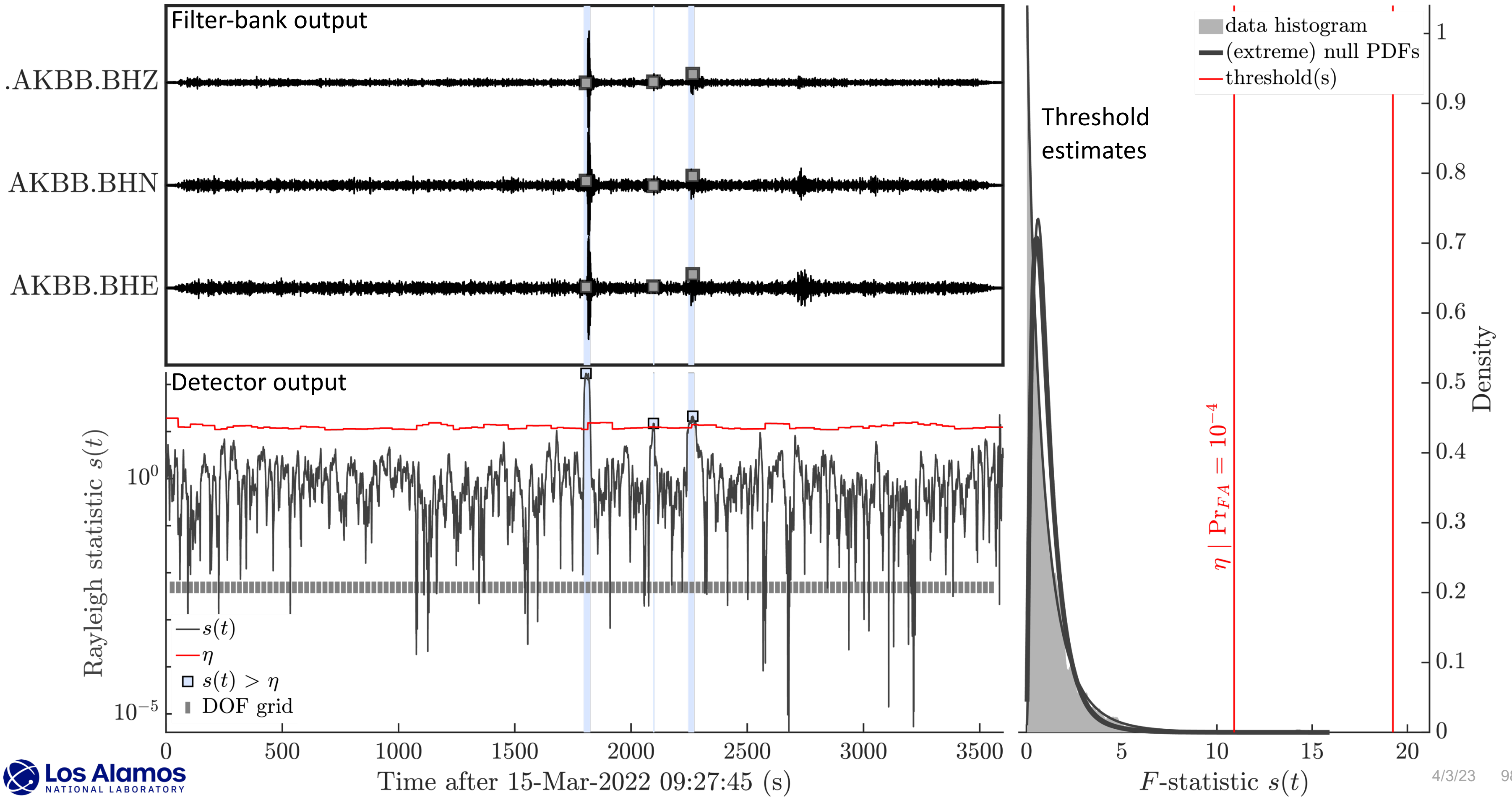


Thresholds depend on density shape. Density shape depends on effective degrees of freedom (DOF). Thresholds therefore depend on DOFs.

4. Hold the first DOF fixed at the Bayesian updated values.
5. Repeat steps 1-3 for the second DOF.
6. Each grid point that ticks | mark now has its own DOF density parameters.
7. Compute a threshold t at each grid point.

Module 8: The Rayleigh Wave Detector Algorithm

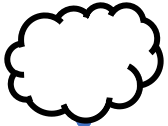
Module 8: The Rayleigh Wave Detector Algorithm (1/1)



Recap and Summary

The Five Steps of Detection (think Scientific Method)

Single modality detection: terminology, concepts



$$\left.\begin{aligned}\mathcal{H}_0\colon\ \boldsymbol{x} &= \text{noise} + \text{background}\\ \mathcal{H}_1\colon\ \boldsymbol{x} &= \text{noise} + \text{background} + \text{signal}\end{aligned}\right\}\ \frac{\widehat{\mathrm{PDF}}(\boldsymbol{x}|\mathcal{H}_1)}{\widehat{\mathrm{PDF}}(\boldsymbol{x}|\mathcal{H}_0)} = s(\boldsymbol{x})\ \underset{\mathcal{H}_0}{\overset{\mathcal{H}_1}{\gtrless}}\ t \qquad \begin{aligned}\mathcal{H}_1 &\rightarrow u = 1\\ \mathcal{H}_0 &\rightarrow u = 0\end{aligned}$$

Source — **Origin of observables**

Data Models — **Form competing hypotheses**; **Gather data / observables**

GLR — **Analyze data/observables**

Detector — **Decision** — **Accept or reject initial hypothesis**

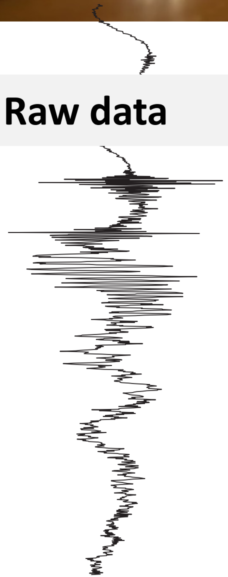
Density processing: estimate distributional parameters that shape density and histograms

Data pre-processing: detrend, taper and bandpass filter (possibly in a filter bank)

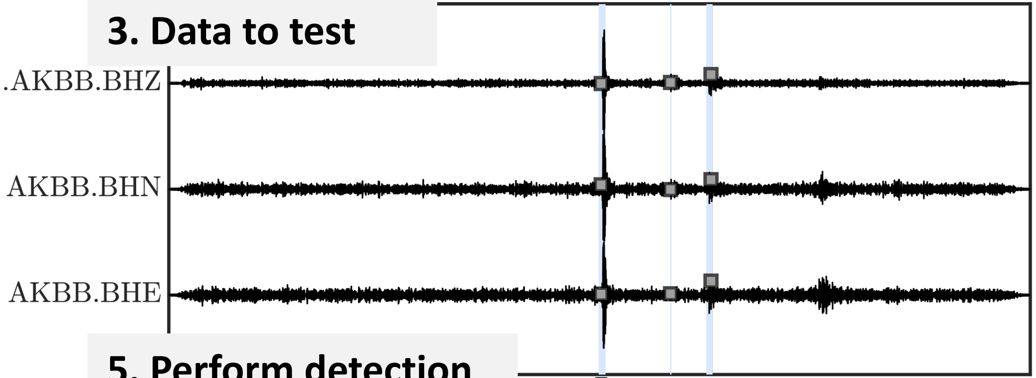
Bottom Line Up Front (Low-Fidelity BLUF): We Did…

1. Event?

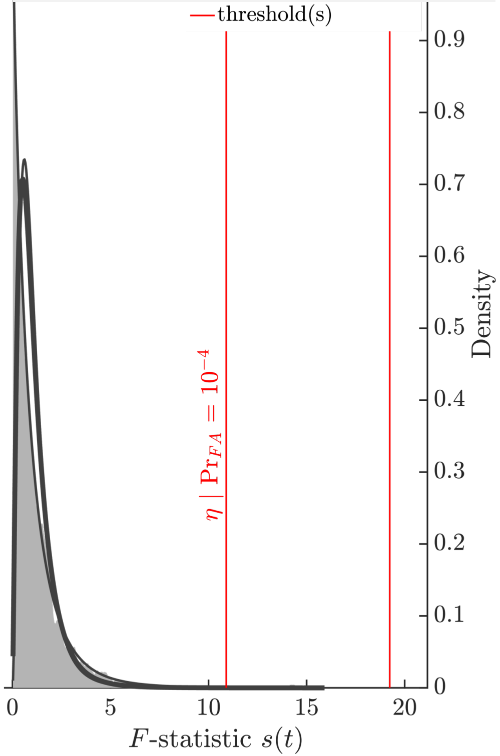
2. Raw data



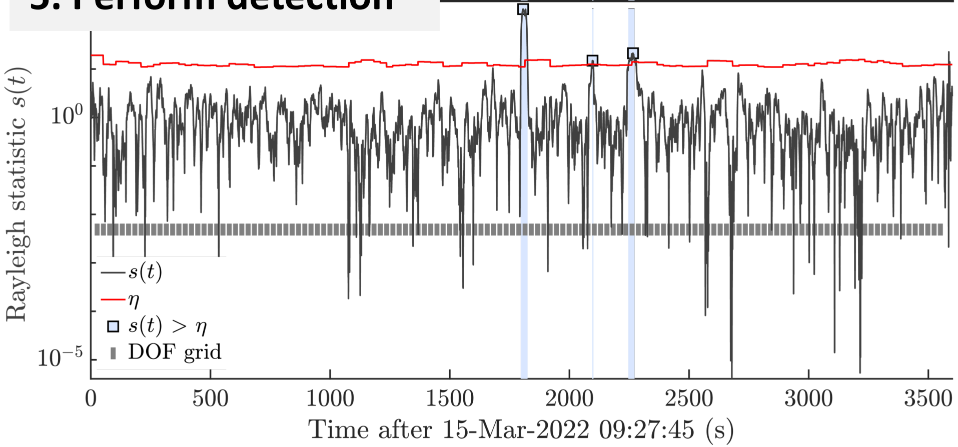
3. Data to test



4. Estimate thresholds



5. Perform detection



Module 9: Homework Exercises (Completed Homework Solutions will be Published Separately)

Module 9: Homework Exercises; Exercise 1 (1/10)

Problem: Consider the Rayleigh wave digital signal detector that Module 4 discussed. Perform the maximization required to compute just the numerator of the GLRT (similar derivations can be found in references elsewhere). Recall:

Data is only noise $\mathcal{H}_0$: $\mathcal{J}[\boldsymbol{x}_Z] = \boldsymbol{n} \sim \mathcal{N}(\boldsymbol{0}, \sigma^2 \boldsymbol{I})$

Data includes Rayleigh wave $\mathcal{H}_1$: $\mathcal{J}[\boldsymbol{x}_Z] = [\boldsymbol{x}_E \quad \boldsymbol{x}_N]\boldsymbol{\theta} + \boldsymbol{n} \sim \mathcal{N}([\boldsymbol{x}_E \quad \boldsymbol{x}_N]\boldsymbol{\theta}, \sigma^2 \boldsymbol{I})$

$$\equiv \boldsymbol{U\theta} + \boldsymbol{n} \sim \mathcal{N}(\boldsymbol{U\theta}, \sigma^2 \boldsymbol{I})$$

$$\mathrm{GLR} = \max_{\sigma^2\ \boldsymbol{\theta}} \left\{ \frac{1}{(2\pi\sigma^2)^{\frac{N}{2}}} \exp\left[-\frac{\|\mathcal{J}[\boldsymbol{x}_Z] - \boldsymbol{U\theta}\|^2}{2\sigma^2} \right] \right\} \Bigg/ \max_{\sigma^2} \left\{ \frac{1}{(2\pi\sigma^2)^{\frac{N}{2}}} \exp\left[-\frac{\|\mathcal{J}[\boldsymbol{x}_Z]\|^2}{2\sigma^2} \right] \right\}$$

Module 9: Homework Exercises; Exercise 2

Problem: Consider again the Rayleigh wave digital signal detector that Module 4 discussed. Note the detection statistic for non-zero amplitude signals. Argue that the distributional form of the detection statistic is central *F* as sample number N becomes much greater than one.

Recall:

$$s(\boldsymbol{x}_Z) = \left(\frac{N-2}{2}\right) \frac{\boldsymbol{x}_Z^{\mathrm{T}} \boldsymbol{U}\left(\boldsymbol{U}^{\mathrm{T}}\boldsymbol{U}\right)^{-1} \boldsymbol{U}^{\mathrm{T}} \boldsymbol{x}_Z}{\boldsymbol{x}_Z^{\mathrm{T}}(\boldsymbol{I} - \boldsymbol{U}(\boldsymbol{U}^{\mathrm{T}}\boldsymbol{U})^{-1}\boldsymbol{U}^{\mathrm{T}})\boldsymbol{x}_Z}, \qquad \text{and:}$$

$$\boldsymbol{U\theta} + \boldsymbol{n} \sim \mathcal{N}(\boldsymbol{U\theta}, \sigma^2 \boldsymbol{I}), \qquad \text{and:}$$

$$\boldsymbol{n} \sim \mathcal{N}(\boldsymbol{0}, \sigma^2 \boldsymbol{I}).$$

Module 9: Homework Exercises; Exercise 3

Problem: Consider again the Rayleigh wave digital signal detector that Module 4 discussed. Write the form of the noncentrality parameter $\lambda = \boldsymbol{\theta}^{\mathrm{T}}\boldsymbol{\theta}/\sigma^2$ in terms of α, A and B.

Module 9: Homework Exercises; Exercise 4

Problem: Consider a scenario: A three channel sensor records a waveform template $[\boldsymbol{u}_E, \boldsymbol{u}_N, \boldsymbol{u}_Z]$. The sensor then rotates an unknown angle. The observer must still detect signals sourced by repeating events that match the waveform shape of the original waveform template. Design a GLRT to detect signals in noise that accommodates this rotation. *Hint: consider the orthogonal Procrustes problem, with $\boldsymbol{Q}$ a rotation matrix:*

Data is only noise $\mathcal{H}_0$: $[\boldsymbol{x}_E, \boldsymbol{x}_N, \boldsymbol{x}_Z] = [\boldsymbol{n}_E, \boldsymbol{n}_N, \boldsymbol{n}_Z]$

Data includes a scaled, rotated copy of the template $\mathcal{H}_1$: $[\boldsymbol{x}_E, \boldsymbol{x}_N, \boldsymbol{x}_Z] = [\boldsymbol{n}_E, \boldsymbol{n}_N, \boldsymbol{n}_Z] + A[\boldsymbol{u}_E, \boldsymbol{u}_N, \boldsymbol{u}_Z]\boldsymbol{Q}$

Module 9: Homework Exercises; Exercise 5

Problem: Consider a binary hypothesis test against a constrained mean vector in which the noise includes an additional, unknown component of variance under the alternative. Both data have an unknown mean vector and a known variance component of σ_0^2. From the GLR, compute the detection statistic. Substantial effort is required to compute its performance. *Hint: The density function for this GLR requires the Lambert function. The hypothesis test is:*

Constrained: $\mathcal{H}_0$: $\mathcal{J}[\boldsymbol{x}_Z] = \boldsymbol{U\theta} + \boldsymbol{n} \sim \mathcal{N}(\boldsymbol{U\theta}, \sigma_0^2 \boldsymbol{I})$ constrained by $[0 \quad 1 \quad 1]\boldsymbol{\theta} = 0.$

Unconstrained: $\mathcal{H}_1$: $\mathcal{J}[\boldsymbol{x}_Z] = \boldsymbol{U\theta} + \boldsymbol{n} \sim \mathcal{N}(\boldsymbol{U\theta}, (\sigma_1^2 + \sigma_0^2)\boldsymbol{I})$

Equation 35 in doi:10.1093/gji/ggab055 provides the solution

Module 9: Homework Exercises; Exercise 6

Problem: Consider the equation for the Baye's estimate of the first noncentrality parameter discussed in Module 6, (8/11). Suppose the density for the Rayleigh statistic uses four data samples on the coarse grid ($\widehat{D1}$ associates to grid point k). Describe the significance of each term in the Bayesian estimate for $D1$. Note C, $s = 2$, $s = 2TBP$, and $\beta(s)$ are not explained or defined, because <u>*you*</u> should describe them by referring to Baye's estimates, the *F*-distribution, the time-bandwidth product, and the Beta distribution.

$$\widehat{D1} = \frac{1}{C}\int_{s=2}^{s=2TBP} \left[f_{D1,\widehat{D2}}(z_{k-4};\mathcal{H}_0)\ldots f_{D1,\widehat{D2}}(z_k;\mathcal{H}_0)\right]\beta(s)ds$$

Extra Slides

Module 3: Form Competing Data Hypothesis [EXTRA]

Data is only noise: $\mathcal{H}_0: [\boldsymbol{x}_E \quad \boldsymbol{x}_N \quad \boldsymbol{x}_Z] = \boxed{[\boldsymbol{n}_E \quad \boldsymbol{n}_N \quad \boldsymbol{n}_Z]}$

Data includes Rayleigh wave $\mathcal{H}_1: [\boldsymbol{x}_E \quad \boldsymbol{x}_N \quad \boldsymbol{x}_Z] = \boxed{[\boldsymbol{n}_E \quad \boldsymbol{n}_N \quad \boldsymbol{n}_Z] + [\boldsymbol{s}_E \quad \boldsymbol{s}_N \quad \boldsymbol{s}_Z]}$

Consider the north to radial conversion: https://service.iris.edu/irisws/rotation/docs/1/help/

Two dimensional rotation

The rotation service uses the following transformation matrix to change the output vectors for 2-D horizontal transformations

$$M_{2D} = \begin{bmatrix} \cos a & \sin a \\ -\sin a & \cos a \end{bmatrix} \quad \begin{bmatrix} R \\ T \end{bmatrix} = M_{2D} \begin{bmatrix} N \\ E \end{bmatrix}$$

where :

- **N** , and **E** represent data from the original (horizontal) orientations
- **R**,and **T** represent the Radial and Transverse components.
- **a** is the azimuth measured clockwise from north